Bernoulli **23**(3), 2017, 1951–1996
DOI: 10.3150/16-BEJ801

Efficient particle-based online smoothing in general hidden Markov models: The PaRIS algorithm

JIMMY OLSSON[*] and JOHAN WESTERBORN[**]

Department of Mathematics, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden.
E-mail: [*]*jimmyol@kth.se;* [**]*johawes@kth.se*

This paper presents a novel algorithm, the particle-based, rapid incremental smoother (PaRIS), for efficient online approximation of smoothed expectations of additive state functionals in general hidden Markov models. The algorithm, which has a linear computational complexity under weak assumptions and very limited memory requirements, is furnished with a number of convergence results, including a central limit theorem. An interesting feature of PaRIS, which samples on-the-fly from the retrospective dynamics induced by the particle filter, is that it requires two or more backward draws per particle in order to cope with degeneracy of the sampled trajectories and to stay numerically stable in the long run with an asymptotic variance that grows only linearly with time.

Keywords: central limit theorem; general hidden Markov models; Hoeffding-type inequality; online estimation; particle filter; particle path degeneracy; sequential Monte Carlo; smoothing

1. Introduction

This paper deals with the problem of state estimation in *general state-space hidden Markov models* (HMMs) using *sequential Monte Carlo* (SMC) *methods* (also known as *particle filters*), and presents a novel online algorithm for the computation of smoothed expectations of *additive state functionals* in models of this sort. The algorithm, which copes with the well-known problem of *particle ancestral path degeneracy* at a computational complexity that is only *linear* in the number of particles, is provided with a rigorous theoretical analysis establishing its convergence and long-term stability as well as a simulation study illustrating its computational efficiency.

Given measurable spaces $(\mathsf{X}, \mathcal{X})$ and $(\mathsf{Y}, \mathcal{Y})$, an HMM is a bivariate stochastic process $\{(X_t, Y_t)\}_{t\in\mathbb{N}}$ (where t will often be referred to as "time" without being necessarily a temporal index) taking its values in the product space $(\mathsf{X} \times \mathsf{Y}, \mathcal{X} \otimes \mathcal{Y})$, where the X-valued marginal process $\{X_t\}_{t\in\mathbb{N}}$ is a *Markov chain* (often referred to as the *state sequence*) which is only partially observed through the Y-valued *observation process* $\{Y_t\}_{t\in\mathbb{N}}$. Conditionally on the unobserved state sequence $\{X_t\}_{t\in\mathbb{N}}$, the observations are assumed to be independent and such that the conditional distribution of each Y_t depends on the corresponding state X_t only. HMMs are nowadays used within a large variety of scientific and engineering disciplines such as econometrics [6], speech recognition [36] and computational biology [27] (the more than 360 references in [2] for the period 1989–2000 gives an idea of the applicability of these models).

Any kind of statistical inference in HMMs involves typically the computation of conditional distributions of unobserved states given observations. Of particular interest are the sequences of *filter distributions*, that is, the conditional distributions of X_t given $Y_{0:t} := (Y_0, \dots, Y_t)$ (this will be our generic notation for vectors), and *smoothing distributions*, that is, the joint conditional distributions of $X_{0:t}$ given $Y_{0:t}$, for $t \in \mathbb{N}$. We will denote these distributions by ϕ_t and $\phi_{0:t|t}$, respectively (precise definitions of these measures are given in Section 2.2). In this paper, we are focusing on the problem of computing, recursively in time, smoothed expectations

$$\phi_{0:t|t} h_t = \int h_t(x_{0:t}) \phi_{0:t|t}(\mathrm{d}x_{0:t}) \qquad (t \in \mathbb{N}), \tag{1.1}$$

for additive functionals h_t of form

$$h_t(x_{0:t}) := \sum_{\ell=0}^{t-1} \tilde{h}_\ell(x_{\ell:\ell+1}) \qquad \big(x_{0:t} \in \mathsf{X}^{t+1}\big). \tag{1.2}$$

Expectations of the form (1.1) appear naturally in the context of parameter estimation using the *maximum-likelihood method*, for example, when computing the *score-function* (the gradient of the log-likelihood function) via the *Fisher identity* or when computing the *intermediate quantity* of the *expectation-maximization* (EM) *algorithm*. Of particular relevance is the situation where the HMM belongs to an *exponential family*. We refer to [5], Sections 10 and 11, for a comprehensive treatment of these matters. Moreover, *online* implementations of EM (see, e.g., [3,29]) require typically such smoothed expectations to be computed in an online fashion. In the case of *marginal smoothing*, the interest lies in computing conditional expectations of some state $X_{\hat{s}}$ given $Y_{0:t}$ for $t \geq \hat{s}$, which can be cast into our framework by letting, in (1.2), $\tilde{h}_\ell = 0$ for $\ell \neq \hat{s}$ and $\tilde{h}_{\hat{s}}(x_{\hat{s}:\hat{s}+1}) = \tilde{h}_{\hat{s}}(x_{\hat{s}})$. Nevertheless, since exact computation of smoothed expectations is possible only in the cases of linear Gaussian HMMs or HMMs with finite state space, we are in general referred to finding approximations of these quantities, and the present paper focuses on the use of SMC-based techniques for this task. A particle filter approximates the flow $\{\phi_t\}_{t \in \mathbb{N}}$ of filter distributions by a sequence of occupation measures associated with samples $\{(\xi_t^i, \omega_t^i)\}_{i=1}^N$, $t \in \mathbb{N}$, of random draws, *particles* (the ξ_t^i's), with associated non-negative importance weights (the ω_t^i's). Particle filters revolve around two operations: a *selection step* duplicating/discarding particles with large/small importance weights, respectively, and a *mutation step* evolving randomly the selected particles in the state space. The first and most basic implementation, the so-called *bootstrap particle filter* [21] (see also [25]), propagates, in the mutation step, the particles according to the dynamics of the hidden Markov chain and selects the same multinomially according to importance weights proportional to the local likelihood of each particle given the current observation. This scheme imposes a dynamics of the particle cloud that resembles closely that of the filter distribution flow. Due to its very strong potential to solve non-linear/non-Gaussian filtering problems, SMC methods have been subject to extensive research during the last two decades, resulting in a broad range of developments and variations of the original scheme; see, for example, [4,5,15,17] and the references therein.

Interestingly, the particle filter provides, as a by-product, approximations also of the joint smoothing distributions in the sense that for each $t \in \mathbb{N}$, the occupation measure associated with the *ancestral lines* of the particles $\{\xi_t^i\}_{i=1}^N$ forms, when the lines are assigned the corresponding weights $\{\omega_t^i\}_{i=1}^N$, an estimate of $\phi_{0:t|t}$. Unfortunately, this *poor man's smoother* (using the terminology of [14]) has a major flaw in that resampling systematically the particles leads to significant depletion of the trajectories and the existence of a random time before which all the ancestor paths coincide. In fact, [24] established, in the case of a compact state space X, a bound on the expected height of the "crown" of the ancestral tree, that is, the expected time distance from the last generation back to the most recent common ancestor, which is proportional to $N \log(N)$ and *uniform* in time. Thus, the ratio of the length of the "crown" to that of the "trunk" tends to zero when time increases, implying that the Monte Carlo approximation obtained through this naive approach will, for long observation records, be based on practically a *single draw*, leading to a severely depleted estimator.

Haplessly, the particle path degeneracy implies that the variance of estimates produced using the poor man's smoother grows *quadratically* with t. To illustrate this phenomenon, let η and h be a probability measure and a measurable function, respectively, on some state space $(\mathsf{X}, \mathcal{X})$, and consider the problem of estimating expectations $m_t = \int h_t(x_{1:t}) \eta^{\otimes t}(\mathrm{d}x_{1:t})$, where $h_t(x_{1:t}) = \sum_{s=1}^t h(x_s)$ and $\eta^{\otimes t} = \eta \otimes \cdots \otimes \eta$ (t times) denotes the product measure on $(\mathsf{X}^t, \mathcal{X}^{\otimes t})$. Assume that $\sigma^2 = \mathrm{Var}_\eta(h) = \int \{h(x) - m_1\}^2 \eta(\mathrm{d}x) < \infty$. For this purpose, we consider two different approaches, namely

(1) The standard Monte Carlo approach, which generates a sample $\{\hat{\xi}_{1:t}^i\}_{i=1}^N$ of i.i.d. random variables from $\eta^{\otimes t}$ and provides the estimator $\hat{m}_t^N = \sum_{i=1}^N h_t(\hat{\xi}_{1:t}^i)/N$. The estimator $\hat{m}_t^N$ is unbiased for all N and has, straightforwardly, the variance $\hat{\sigma}_t^N = \sigma^2 t/N$, which is *linear* in t.

(2) A selection and sampling approach with genealogical tracing, which resembles closely the poor man's smoother and proceeds recursively as follows: given a sample $\{\check{\xi}_{1:s}^i\}_{i=1}^N$ of random variables such that the estimator $\check{m}_s^N = \sum_{i=1}^N h_s(\check{\xi}_{1:s}^i)/N$ approximates m_s, (i) draw a set $\{I^i\}_{i=1}^N$ of i.i.d. indices from the uniform distribution on $\{1, 2, \ldots, N\}$, (ii) draw a set $\{\check{\xi}_{s+1}\}_{i=1}^N$ of i.i.d. variables from η and (iii) set $\check{\xi}_{1:s+1}^i = (\check{\xi}_{1:s}^{I^i}, \check{\xi}_{s+1}^i)$. In other words, at each step of this procedure, each draw $\check{\xi}_{1:s+1}^i$ is formed by selecting an ancestor with uniform probability among $\{\check{\xi}_{1:s}^i\}_{i=1}^N$ and extending the same by a draw from η. The procedure, which is initialized by letting $\{\check{\xi}_1^i\}_{i=1}^N$ be i.i.d. draws from η, is repeated recursively for $s \in \{1, 2, \ldots, t-1\}$. For each s, the estimator $\check{m}_s^N$ is unbiased w.r.t. m_s for all N. Moreover, the variance can be shown to satisfy the recursion

$$\check{\sigma}_{s+1,N}^2 = \frac{1}{N} \mathrm{Var}_{\eta^{\otimes(s+1)}}(h_{s+1}) + \check{\sigma}_{s,N}^2 \left(1 - \frac{1}{N}\right) = \frac{1}{N} \sigma^2 (s+1) + \check{\sigma}_{s,N}^2 \left(1 - \frac{1}{N}\right) \quad (1.3)$$

(with $\check{\sigma}_{1,N}^2 = \sigma^2/N$) implying that

$$\check{\sigma}_{t,N}^2 = \sigma^2 \frac{1}{N} \sum_{s=1}^t s \left(1 - \frac{1}{N}\right)^{t-s}. \quad (1.4)$$

When N tends to infinity, the sum in (1.4) tends to a *quadratic* function in t; indeed, entering the asymptotic regime, the normalized estimator can be shown to satisfy the central limit theorem (CLT)

$$\sqrt{N}\left(\check{m}_t^N - m_t\right) \xrightarrow{\mathcal{D}} \check{\sigma}_{t,\infty} Z,$$

where Z has standard Gaussian distribution and the asymptotic variance

$$\check{\sigma}_{t,\infty}^2 = \sigma^2 \sum_{s=1}^{t} s = \sigma^2 \frac{t(t+1)}{2}$$

is indeed quadratic in t.

As clear from (1.3), the quadratic variance growth of the second procedure comes from the fact that the multinomial selection step adds, at iteration s, when followed by the subsequent sampling step, an $\mathcal{O}(s)$ variance term equal to the variance of the state functional under the target distribution restricted to the first $s+1$ components of the path space. As the same principle applies generally to path-space applications of the standard SMC method with multinomial resampling (see [12], Theorem 4, for a single-step analysis in the general case and treatments of alternative resampling strategies), this toy example pinpoints the crux of the poor man's smoother from a variance point of view; see also [35] for a discussion. On the contrary, as the components of the target measure $\eta^{\otimes t}$ are independent in this simple model, the naive Monte Carlo sampler in Procedure 1, where the draws do not interact, attains the optimal linear variance growth. However, this does not hold in the general case, involving typically dependent components and non-uniform particle weights, where particle interaction is crucial for preserving the numerical stability of the algorithm.

1.1. Previous work

In the case of additive state functionals, it is possible to cope partly with the degeneracy problem described above by means of a *fixed-lag smoothing* technique [26,30,32]. This approach avoids the particle path degeneracy by "localizing" the smoothing of a certain state around observations that are only significantly statistically dependent of the state in question and discarding remote and weakly influential observations, that is, subsequent observations located at a time distance from the state exceeding a lag chosen by the user. The method is expected to work well if the mixing properties of the model allow the lag to be smaller than the length of the "crown" of the ancestral tree. Still, such truncation introduces a mixing-dependent bias, and designing the size of the lag is thus a non-trivial task.

A completely different way of approaching the problem goes via the so-called *forward-filtering backward-smoothing decomposition*, which is based on the fact that the latent process still satisfies the Markov property when evolving backward in time and conditionally on the observations. Consequently, each smoothing measure $\phi_{0:t|t}$ can be represented as the joint law of this inhomogeneous backward chain with initial distribution given by the corresponding filter ϕ_t. Since the transition kernels of the backward chain depend on the filter distributions, which may be estimated efficiently by a particle filter, a particle-based approximation of the smoothing distribution can thus be naturally obtained by running, in a prefatory filtering pass, the particle filter up to time t (if $\phi_{0:t|t}$ is the distribution of interest) and, in a backward pass, forming particle-based estimates of the backward kernels (and consequently the smoothing distribution) by modifying the particle weights computed in the forward pass. This scheme, which avoids completely the path degeneracy problem at the cost of a rather significant computational complexity, is referred to as the *forward-filtering backward smoothing* (FFBSm) *algorithm* [16,23,25]. As an alternative, the *forward-filtering backward simulation* (FFBSi) *algorithm* [20] generates, in order to reduce the computational overhead of FFBSm, trajectories being approximately distributed according to the smoothing distribution by *simulating* transitions according to the backward dynamics induced by the particle filter approximations produced by the forward pass; as a consequence, FFBSm can be viewed as a Rao–Blackwellized version of FFBSi. These two algorithms correspond directly to the *Rauch–Tung–Striebel smoother* [37] for linear Gaussian HMMs or the *Baum–Welch algorithm* [1] for HMMs with finite state space. FFBSm and FFBSi were analyzed theoretically in [11] (see also [9]), which provides exponential concentration inequalities as well as CLTs for these algorithms. Since each backward draw of FFBSi requires a normalizing constant with N terms to be computed, the overall complexity of the algorithm is $\mathcal{O}(N^2)$. Under the mild assumption that the transition density of the latent chain is uniformly bounded, this complexity can be reduced to $\mathcal{O}(N)$ by means of simple accept–reject approach. The latter technique, which was found in [11], will play a key role also in the development of the present paper. Since the Markov transition kernels of the backward chain depend on the filter distributions, the FFBSm and FFBSi algorithms require in general *batch mode* processing of the observations. This is the case also for the $\mathcal{O}(N)$ smoother proposed in [19], which is based on the *two-filter representation* of each marginal smoothing distribution.

When the objective consists in *online* smoothing of additive state functionals (1.2), *recursive* approximation of the forward-filtering backward-smoothing decomposition can be achieved by introducing the auxiliary statistics

$$\mathbf{T}_t h_t(x_t) = \mathbb{E}\big[h_t(X_{0:t}) \mid X_t = x_t, Y_{0:t}\big] \qquad (t \in \mathbb{N}, x_t \in \mathsf{X}),$$

where $\mathbb{E}$ denotes expectation associated with the law of the canonical version of the HMM (more precisely, in the previous expression $\mathbf{T}_t$ is a normalized transition kernel which will be defined in Section 2.2). This auxiliary statistic can be updated online according to

$$\mathbf{T}_{t+1} h_{t+1}(x_{t+1}) = \mathbb{E}\big[\mathbf{T}_t h_t(X_t) + \tilde{h}_t(X_t, x_{t+1}) \mid X_{t+1} = x_{t+1}, Y_{0:t}\big] \\ (t \in \mathbb{N}, x_{t+1} \in \mathsf{X}); \tag{1.5}$$

see [3,9,29]. In this recursive formula, the expectation is taken under the backward kernel describing the conditional distribution of X_t given X_{t+1} and $Y_{0:t}$. On the basis of the auxiliary statistics, each smoothed additive functional may be computed as

$$\phi_{0:t|t} h_t = \int \mathbf{T}_t h_t(x_t) \phi_t(\mathrm{d}x_t) \qquad (t \in \mathbb{N}).$$

Following [9], a particle representation of the recursion (1.5) is naturally formed using the estimates of the retrospective dynamics provided by the FFBSm algorithm. Interestingly, this yields a procedure that estimates, as new observations become available, the smoothing distribution flow in a forward-only manner while avoiding completely any problems of particle path degeneracy.

However, since the method requires the normalizations of the backward kernels to be computed for each forward particle, the overall complexity of this algorithm is again $\mathcal{O}(N^2)$, which is unrealistic for large particle sample sizes.

1.2. Our approach

Our novel algorithm, which we will refer to as the *particle-based, rapid incremental smoother* (PaRIS), is, similarly to the forward-only implementation of FFBSm proposed in [9], based on (1.5) and can be viewed as an adaptation of the FFBSi algorithm to this recursion. It also shares some similarities with the ancestor sampling approach within the framework of particle Gibbs sampling [28]. Appealingly, we are able to adopt the accept-technique proposed by [11], yielding a fast algorithm with $\mathcal{O}(N)$ complexity. PaRIS differs from the forward-only implementation of FFBSm in the way the update (1.5) of the auxiliary function is implemented; more specifically, instead of computing each subsequent auxiliary statistic as the expected sum of the previous statistic and the incremental term under the retrospective dynamics induced by the particle filter, PaRIS *simulates* $\tilde{N}$ such sums using the backward kernel and updates each statistic by taking the *sample mean* of these draws. Thus, as for the FFBSi algorithm, forward-only FFBSm can be viewed as a Rao–Blackwellization of PaRIS. Interestingly, the design of the sample size $\tilde{N}$ is ultimately critical, as the naive choice $\tilde{N} = 1$ leads to a degeneracy phenomenon that resembles closely that of the Poor man's smoother and, consequently, a variance that grows quadratically with t; on the other hand, for all $\tilde{N} \geq 2$ the algorithm stays numerically stable in the long run with a *linearly* increasing variance. The main objective of the present paper is to investigate theoretically this phase transition by, first, deriving, via a non-asymptotic Hoeffding-type inequality, the asymptotic (as N tends to infinity) variance of the Monte Carlo estimates produced by the algorithm (which is highly non-trivial due to the complex dependence structures induced by the backward simulation) and, second, verifying that this asymptotic variance is, for any $\tilde{N} \geq 2$, of order $\mathcal{O}(t)$ and $\mathcal{O}(1)$ in the cases of joint smoothing and marginal smoothing, respectively. The authors are not aware of any similar analysis in the SMC literature. The stability results are obtained under strong mixing assumptions that are standard in the literature of SMC analysis (see, e.g., [5,8,10]). Also the numerical performance of algorithm is investigated in a simulation study, comprising a linear Gaussian state space model (for which any quantity of interest may be computed exactly using the Rauch–Tung–Striebel smoother) and a stochastic volatility model.

We finally point out that the PaRIS algorithm was outlined by us in the conference note [33] without any theoretical support; in the present paper we are able to confirm, through a rigorous theoretical analysis, the conjectures made in the note in question concerning the stability properties of the algorithm.

To sum up, the smoothing algorithm we propose:

- is computationally very efficient and easy to implement,
- does not suffer from particle lineage degeneracy,
- allows the observed data of the HMM to be processed online with minimal memory requirements, and
- is furnished with rigorous theoretical results describing the convergence and numeric stability of the same.

1.3. Outline

After having introduced some kernel notation, HMMs, and the smoothing problem in Section 2, we describe carefully, in Section 2.3, particle filters, FFBSm (and its forward-only implementation), and FFBSi. Section 3.1 contains the derivation of our novel algorithm as well as some discussion of the choice of the design parameter $\tilde{N}$. Our theoretical results are presented in Section 3.2, including a Hoeffding-type inequality (Theorem 1) and a CLT (Theorem 3). Section 3.2.3 is devoted to the numerical stability of PaRIS in the case $\tilde{N} \geq 2$, and Theorem 8 and Theorem 9 provide variance bounds in the cases of joint and marginal smoothing, respectively. In Section 4, we test numerically the algorithm and some conclusions are drawn in Section 5. Finally, Appendix A and Appendix B provide all proofs and some technical results, respectively.

2. Preliminiaries

2.1. Notation

Before going into the details concerning HMMs and particle filters we introduce some notation. For any measurable space $(\mathsf{X}, \mathcal{X})$, where $\mathcal{X}$ is a countably generated σ-algebra, we denote by $\mathsf{F}_\mathrm{b}(\mathcal{X})$ the set of bounded $\mathcal{X}/\mathcal{B}(\mathbb{R})$-measurable functions on X. For any $h \in \mathsf{F}_\mathrm{b}(\mathcal{X})$, we let $\|h\|_\infty := \sup_{x \in \mathsf{X}} |h(x)|$ and $\operatorname{osc}(h) := \sup_{(x,x') \in \mathsf{X}^2} |h(x) - h(x')|$ denote the sup and oscillator norms of h, respectively. Let $\mathsf{M}(\mathcal{X})$ be the set of σ-finite measures on $(\mathsf{X}, \mathcal{X})$ and $\mathsf{M}_1(\mathcal{X}) \subset \mathsf{M}(\mathcal{X})$ the probability measures. Given $n \in \mathbb{N}$, we will denote product sets and product σ-fields by $\mathsf{X}^n := \mathsf{X} \times \cdots \times \mathsf{X}$ and $\mathcal{X}^n := \mathcal{X} \otimes \cdots \otimes \mathcal{X}$ (n times), respectively. For real numbers and integers, we define the sets $\mathbb{R}_+ := [0, \infty)$, $\mathbb{R}_+^* := (0, \infty)$, $\mathbb{N} := \{0, 1, 2, \dots\}$, and $\mathbb{N}^* := \{1, 2, 3, \dots\}$. For any quantities $\{a_\ell\}_{\ell=m}^n$, we denote vectors as $a_{m:n} := (a_m, \dots, a_n)$ and for any $(m, n) \in \mathbb{N}^2$ such that $m \leq n$ we denote $[\![m, n]\!] := \{m, m+1, \dots, n\}$. The cardinality of a set S is denoted by $\#\mathsf{S}$.

An unnormalized transition kernel $\mathbf{K}$ from $(\mathsf{X}, \mathcal{X})$ to $(\mathsf{Y}, \mathcal{Y})$ induces two integral operators, one acting on functions and the other on measures. More specifically, let $h \in \mathsf{F}_\mathrm{b}(\mathcal{X} \otimes \mathcal{Y})$ and $\nu \in \mathsf{M}(\mathcal{X})$, and define the measurable function

$$\mathbf{K}h : \mathsf{X} \ni x \mapsto \int h(x, y) \mathbf{K}(x, \mathrm{d}y),$$

and the measure

$$\nu\mathbf{K} : \mathcal{Y} \ni \mathsf{A} \mapsto \int \mathbf{K}(x, \mathsf{A}) \nu(\mathrm{d}x),$$

whenever these quantities are well-defined. Moreover, let $\mathbf{K}$ be defined as above and let $\mathbf{L}$ be another unnormalized transition kernel from $(\mathsf{Y}, \mathcal{Y})$ to a third measurable space $(\mathsf{Z}, \mathcal{Z})$; we then define two different *products* of $\mathbf{K}$ and $\mathbf{L}$, namely

$$\mathbf{KL} : \mathsf{X} \times \mathcal{Z} \ni (x, \mathsf{A}) \mapsto \int \mathbf{K}(x, \mathrm{d}y) \mathbf{L}(y, \mathsf{A})$$

and

$$\mathbf{K} \otimes \mathbf{L} : \mathsf{X} \times (\mathcal{Y} \otimes \mathcal{Z}) \ni (x, \mathsf{A}) \mapsto \int \mathbb{1}_{\mathsf{A}}(y, z)\mathbf{K}(x, \mathrm{d}y)\mathbf{L}(y, \mathrm{d}z),$$

whenever these are well-defined. Note that the previous products form new transition kernels from $(\mathsf{X}, \mathcal{X})$ to $(\mathsf{Z}, \mathcal{Z})$ and from $(\mathsf{X}, \mathcal{X})$ to $(\mathsf{Y} \times \mathsf{Z}, \mathcal{Y} \otimes \mathcal{Z})$, respectively. We also define the $\otimes$-product of a kernel $\mathbf{K}$ and a measure $\nu \in \mathsf{M}(\mathcal{X})$ as the new measure

$$\nu \otimes \mathbf{K} : \mathcal{X} \otimes \mathcal{Y} \ni \mathsf{A} \mapsto \int \mathbb{1}_{\mathsf{A}}(x, y)\mathbf{K}(x, \mathrm{d}y)\nu(\mathrm{d}x).$$

The concept of *reverse kernels* will be of importance in the coming developments. For a kernel $\mathbf{K}$ from $(\mathsf{X}, \mathcal{X})$ to $(\mathsf{Y}, \mathcal{Y})$ and a probability measure $\eta \in \mathsf{M}_1(\mathcal{X})$, the reverse kernel $\overleftarrow{\mathbf{K}}_\eta$ associated with $(\eta, \mathbf{K})$ is a transition kernel from $(\mathsf{Y}, \mathcal{Y})$ to $(\mathsf{X}, \mathcal{X})$ satisfying, for all $h \in \mathsf{F}_{\mathrm{b}}(\mathcal{X} \otimes \mathcal{Y})$,

$$(\eta \otimes \mathbf{K})h = (\eta\mathbf{K}) \otimes \overleftarrow{\mathbf{K}}_\eta h.$$

A reverse kernel does not always exist; however, if $\mathbf{K}$ has a transition density κ with respect to some reference measure in $\mathsf{M}(\mathcal{Y})$, then $\overleftarrow{\mathbf{K}}_\eta$ exists and is given by

$$\overleftarrow{\mathbf{K}}_\eta h(x) := \frac{\int h(\tilde{x})\kappa(\tilde{x}, x)\eta(\mathrm{d}\tilde{x})}{\int \kappa(\tilde{x}, x)\eta(\mathrm{d}\tilde{x})} \qquad \bigl(h \in \mathsf{F}_{\mathrm{b}}(\mathcal{X}), x \in \mathsf{X}\bigr) \tag{2.1}$$

(see [5], Section 2.1, for details).

Finally, for any kernel $\mathbf{K}$ and any bounded measurable function h we write $\mathbf{K}^2 h := (\mathbf{K}h)^2$ and $\mathbf{K}h^2 := \mathbf{K}(h^2)$. Similar notation will be used for measures.

2.2. Hidden Markov models

Let $(\mathsf{X}, \mathcal{X})$ and $(\mathsf{Y}, \mathcal{Y})$ be some measurable spaces, $\mathbf{Q} : \mathsf{X} \times \mathcal{X} \to [0, 1]$ and $\mathbf{G} : \mathsf{X} \times \mathcal{Y} \to [0, 1]$ some Markov transition kernels, and $\chi \in \mathsf{M}_1(\mathcal{X})$. We define an HMM as the canonical version of the bivariate Markov chain $\{(X_t, Y_t)\}_{t \in \mathbb{N}}$ having transition kernel

$$\mathsf{X} \times \mathsf{Y} \times (\mathcal{X} \otimes \mathcal{Y}) : \bigl((x, y), \mathsf{A}\bigr) \mapsto \mathbf{Q} \otimes \mathbf{G}(x, \mathsf{A}) \tag{2.2}$$

and initial distribution $\chi \otimes \mathbf{G}$. The state process $\{X_t\}_{t \in \mathbb{N}}$ is assumed to be only partially observed through the observations process $\{Y_t\}_{t \in \mathbb{N}}$. The dynamics (2.2) imply that (we refer to [5], Section 2.2, for details)

(i) the state sequence $\{X_t\}_{t \in \mathbb{N}}$ is a Markov chain with transition kernel $\mathbf{Q}$ and initial distribution χ,
(ii) the observations are, conditionally on the states, independent and such that the conditional distribution of each Y_t depends on the corresponding X_t only and is given by the *emission distribution* $\mathbf{G}(X_t, \cdot)$.

We will throughout the paper assume that $\mathbf{G}$ admits a density g (referred to as the *emission density*) with respect to some reference measure $\nu \in \mathsf{M}(\mathcal{Y})$, that is,

$$\mathbf{G}h(x) = \int h(y)g(x,y)\nu(\mathrm{d}y) \qquad \bigl(x \in \mathsf{X}, h \in \mathsf{F}_{\mathrm{b}}(\mathcal{Y})\bigr).$$

In the following, we assume that we are given a distinguished sequence $\{y_t\}_{t\in\mathbb{N}}$ of observations of $\{Y_t\}_{t\in\mathbb{N}}$, and will in general omit the dependence on these observations from the notation. Thus, define $g_t(x) := g(x, y_t)$, $x \in \mathsf{X}$. For any $(s, s', t) \in \mathbb{N}^3$ such that $0 \le s \le s' \le t$ we denote by $\phi_{s:s'|t}$ the conditional distribution (posterior) of $X_{s:s'}$ given the observations $Y_{0:t} = y_{0:t}$. This distribution may be expressed as

$$\phi_{s:s'|t}h = \frac{\int \cdots \int h(x_{s:s'})g_0(x_0)\chi(\mathrm{d}x_0)\prod_{\ell=0}^{t-1} g_{\ell+1}(x_{\ell+1})\mathbf{Q}(x_\ell, \mathrm{d}x_{\ell+1})}{\int \cdots \int g_0(x_0)\chi(\mathrm{d}x_0)\prod_{\ell=0}^{t-1} g_{\ell+1}(x_{\ell+1})\mathbf{Q}(x_\ell, \mathrm{d}x_{\ell+1})} \quad \bigl(h \in \mathsf{F}_{\mathrm{b}}\bigl(\mathcal{X}^{s'-s+1}\bigr)\bigr) \tag{2.3}$$

(assuming that the denominator is non-zero). If $s = s' = t$, we let ϕ_t be shorthand for $\phi_{t|t}$, that is, the filter distribution at time t. If $s = 0$ and $s' = t$, then $\phi_{0:t|t}$ is the joint smoothing distribution. For $t \in \mathbb{N}$, define the unnormalized transition kernels

$$\mathbf{L}_t h(x) := \mathbf{Q}(g_{t+1}h)(x) \qquad \bigl(x \in \mathsf{X}, h \in \mathsf{F}_{\mathrm{b}}(\mathcal{X})\bigr),$$

with the convention that $\mathbf{L}_s\mathbf{L}_t \equiv \mathrm{id}$ whenever $s > t$. In addition, we let $\mathbf{L}_{-1}$ be the Boltzmann multiplicative operator associated with g_0, that is, $\mathbf{L}_{-1}h(x) \equiv g_0(x)h(x)$ for all $h \in \mathsf{F}_{\mathrm{b}}(\mathcal{X})$ and $x \in \mathsf{X}$. By combining this notation with (2.3), we may express each filter distribution as

$$\phi_t = \frac{\chi\mathbf{L}_{-1}\cdots\mathbf{L}_{t-1}}{\chi\mathbf{L}_{-1}\cdots\mathbf{L}_{t-1}\mathbb{1}_{\mathsf{X}}} \qquad (t \in \mathbb{N}),$$

which implies immediately the *filter recursion*

$$\phi_{t+1} = \frac{\phi_t\mathbf{L}_t}{\phi_t\mathbf{L}_t\mathbb{1}_{\mathsf{X}}} \qquad (t \in \mathbb{N}). \tag{2.4}$$

In the following, we will often deal with sums and products of functions with possibly different arguments. Since these functions will be defined on products of X, we will, when needed, with a slight abuse of notation, let subscripts define the domain and the values of such sums and products. For instance, $f_t\tilde{f}_t : \mathsf{X} \ni x_t \mapsto f_t(x_t)\tilde{f}_t(x_t)$ while $f_t + \tilde{f}_{t+1} : \mathsf{X}^2 \ni (x_t, x_{t+1}) \mapsto f_t(x_t) + \tilde{f}_{t+1}(x_{t+1})$.

We will for simplicity assume that the HMM is *fully dominated*, that is, that also $\mathbf{Q}$ admits a transition density q with respect to some reference measure $\mu \in \mathsf{M}(\mathcal{X})$. In this case, the reverse kernel $\overleftarrow{\mathbf{Q}}_\eta$ of $\mathbf{Q}$ with respect to any $\eta \in \mathsf{M}_1(\mathcal{X})$ is well-defined and specified by (2.1) (with $\kappa = q$). It may be shown (see, e.g., [5], Proposition 3.3.6) that the state process has still the Markov property when evolving conditionally on $Y_{0:t} = y_{0:t}$ in the time-reversed direction; moreover, the distribution of X_s given X_{s+1} and $Y_{0:t} = y_{0:t}$ is, for any $s \le t$, given by $\overleftarrow{\mathbf{Q}}_{\phi_s}$, which is referred to as the *backward kernel at time s*. Consequently, we may express each joint smoothing distribution $\phi_{0:t|t}$ as

$$\phi_{0:t|t} = \phi_t \mathbf{T}_t, \tag{2.5}$$

where we have defined the kernels

$$\mathbf{T}_t := \begin{cases} \overleftarrow{\mathbf{Q}}_{\phi_{t-1}} \otimes \overleftarrow{\mathbf{Q}}_{\phi_{t-2}} \otimes \cdots \otimes \overleftarrow{\mathbf{Q}}_{\phi_0} & \text{for } t \in \mathbb{N}^*, \\ \text{id} & \text{for } t = 0 \end{cases}$$

(which were also introduced in (1.5)).

As discussed in the Introduction, the aim of this paper is, given a sequence $\{\tilde{h}_t\}_{t\in\mathbb{N}}$ of terms, to estimate the sequence $\{\phi_{0:t|t} h_t\}_{t\in\mathbb{N}}$, where each h_t is given by (1.2). By convention, $h_0 \equiv 0$ (implying, e.g., that $\mathbf{T}_0 h_0 = 0$). Using (2.5), each quantity of interest may be expressed as $\phi_t \mathbf{T}_t h_t = \phi_{0:t|t} h_t$. In addition, note that $\{\mathbf{T}_t h_t\}_{t\in\mathbb{N}}$ may be expressed recursively as

$$\mathbf{T}_{t+1} h_{t+1} = \overleftarrow{\mathbf{Q}}_{\phi_t} (\mathbf{T}_t h_t + \tilde{h}_t), \tag{2.6}$$

a formula that will play a key role in the coming developments.

Finally, define, for $(s, t) \in \mathbb{N}^2$ such that $s \le t$, the *retro-prospective kernels*

$$\mathbf{D}_{s,t} h(x_s) := \iint h(x_{0:t}) \mathbf{T}_s(x_s, \mathrm{d}x_{0:s-1}) \mathbf{L}_s \cdots \mathbf{L}_{t-1}(x_s, \mathrm{d}x_{s+1:t}),$$
$$\tilde{\mathbf{D}}_{s,t} h(x_s) := \mathbf{D}_{s,t}(h - \phi_{0:t|t} h)(x_s) \qquad \left(x_s \in \mathsf{X}, h \in \mathsf{F}_\mathrm{b}\left(\mathcal{X}^{t+1}\right)\right)$$

operating simultaneously in the backward and forward directions. Note that the only difference between $\mathbf{D}_{s,t}$ and $\tilde{\mathbf{D}}_{s,t}$ is that the latter is centralized around the joint smoothing distribution.

2.3. Particle-based smoothing in HMMs

2.3.1. *The bootstrap particle filter*

In the following, we assume that all random variables are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The bootstrap particle filter updates sequentially in time a set of particles and associated weights in order to approximate the filter distribution flow $\{\phi_t\}_{t\in\mathbb{N}}$ given the sequence $\{y_t\}_{t\in\mathbb{N}}$ of observations. Assume that we have at hand a particle sample $\{(\omega_t^i, \xi_t^i)\}_{i=1}^N$ approximating the filter distribution ϕ_t in the sense that for all $h \in \mathsf{F}_\mathrm{b}(\mathcal{X})$,

$$\phi_t^N h = \sum_{i=1}^N \frac{\omega_t^i}{\Omega_t} h\left(\xi_t^i\right) \overset{N\to\infty}{\simeq} \phi_t h, \tag{2.7}$$

where $\Omega_t := \sum_{i=1}^N \omega_t^i$ denotes the weight sum. To form a weighted particle sample $\{(\omega_{t+1}^i, \xi_{t+1}^i)\}_{i=1}^N$ targeting the subsequent filter ϕ_{t+1}, we simply plug in the approximation ϕ_t^N into the filter recursion (2.4), yielding the approximation of ϕ_{t+1} by a mixture distribution proportional

Algorithm 1 Bootstrap particle filter

Require: A weighted particle sample $\{(\xi_t^i, \omega_t^i)\}_{i=1}^N$ targeting ϕ_t.
1: **for** $i = 1 \to N$ **do**
2: $I_{t+1}^i \sim \mathsf{Pr}(\{\omega_t^\ell\}_{\ell=1}^N)$;
3: draw $\xi_{t+1}^i \sim \mathbf{Q}(\xi_t^{I_{t+1}^i}, \cdot)$;
4: set $\omega_{t+1}^i \leftarrow g_{t+1}(\xi_{t+1}^i)$;
5: **end for**
6: **return** $\{(\xi_{t+1}^i, \omega_{t+1}^i)\}_{i=1}^N$.

to $\sum_{i=1}^N \omega_t^i \mathbf{L}_t(\xi_t^i, \cdot)$, and aim at updating the particle cloud by sampling from this mixture. However, since $\mathbf{L}_t$ is generally intractable, we augment the space by the index i and apply importance sampling from the extended distribution proportional to $\omega_t^i \mathbf{L}_t(\xi_t^i, \cdot)$ using the distribution proportional to $\omega_t^i \mathbf{Q}(\xi_t^i, \cdot)$ as instrumental distribution. This yields a sampling schedule comprising two operations: selection and mutation. In the selection step, a set $\{I_{t+1}^i\}_{i=1}^N$ of indices are drawn multinomially according to probabilities proportional to $\{\omega_t^i\}_{i=1}^N$. After this, the mutation step propagates the particles forward according to the dynamics of the state process and assigns the mutated particles importance weights given by the emission density, that is, for all $i \in [\![1, N]\!]$,

$$\xi_{t+1}^i \sim \mathbf{Q}\big(\xi_t^{I_{t+1}^i}, \cdot\big),$$
$$\omega_{t+1}^i = g_{t+1}\big(\xi_{t+1}^i\big).$$

The algorithm, which is the standard bootstrap particle filter presented in [21], is initialized by drawing $\{\xi_0^i\}_{i=1}^N \sim \chi^{\otimes N}$ and letting $\omega_0^i = g_0(\xi_0^i)$ for all $i \in [\![1, N]\!]$. In this basic scheme, which is summarized in Algorithm 1, the information provided by the most current observation y_{t+1} enters the algorithm via the importance weights only. However, instead of moving the particles "blindly" according to the latent dynamics $\mathbf{Q}$, it is, in order to direct the particle swarm toward regions of the state space with large posterior probability, possible to increase the influence of the last observation on the mutation moves as well as the selection mechanism step via the framework of *auxiliary particle filters* [34]. Even though all the results of the present paper can be extended straightforwardly to auxiliary particle filters, we have chosen to limit the presentation to bootstrap-type particle filters only for clarity.

In Algorithm 1, $\mathsf{Pr}(\{\omega_t^\ell\}_{\ell=1}^N)$ denotes the categorical distribution induced by the probabilities $\{\omega_t^\ell / \Omega_t\}_{\ell=1}^N$. We will express Algorithm 1 in a compact form by writing

$$\text{“}\big\{\big(\xi_{t+1}^i, \omega_{t+1}^i\big)\big\}_{i=1}^N \leftarrow \mathsf{PF}\big(\big\{\big(\xi_t^i, \omega_t^i\big)\big\}_{i=1}^N\big)\text{”}.$$

2.3.2. *Forward-filtering backward-smoothing* (*FFBSm*)

As discussed in Introduction, the bootstrap filter may also be used for smoothing, as the weighted occupation measures associated with the genealogical trees of the particle samples generated by the algorithm form consistent estimates of the joint smoothing distributions. A way of detouring the particle path degeneracy of this poor man's smoother goes via the backward decomposition (2.5), granted that we are able to approximate each kernel $\overleftarrow{\mathbf{Q}}_{\phi_s}$, $s \in \mathbb{N}$. However, considering instead the reverse kernel associated with the particle filter ϕ_s^N, yields, via (2.1), the particle approximations

$$\overleftarrow{\mathbf{Q}}_{\phi_s^N} h(x) = \sum_{i=1}^N \frac{\omega_s^i q(\xi_s^i, x)}{\sum_{\ell=1}^N \omega_s^\ell q(\xi_s^\ell, x)} h(\xi_s^i) \qquad \big(x \in \mathsf{X}, h \in \mathsf{F}_\mathrm{b}(\mathcal{X})\big). \tag{2.8}$$

FFBSm consists in simply inserting these approximations into (2.5), that is, approximating, for $h \in \mathsf{F}_\mathrm{b}(\mathcal{X}^{t+1})$, $\phi_{0:t|t} h$ by

$$\phi_{0:t|t}^N h := \sum_{i_0=1}^N \cdots \sum_{i_t=1}^N \left(\prod_{s=0}^{t-1} \frac{\omega_s^{i_s} q(\xi_s^{i_s}, \xi_{s+1}^{i_{s+1}})}{\sum_{\ell=1}^N \omega_s^\ell q(\xi_s^\ell, \xi_{s+1}^{i_{s+1}})} \right) \frac{\omega_t^{i_t}}{\Omega_t} h(\xi_0^{i_0}, \ldots, \xi_t^{i_t}). \tag{2.9}$$

For general objective functions h, this occupation measure is impractical as the cardinality of its support grows geometrically fast with time. In the case where the objective function h is of additive form (1.2), the computational complexity is still quadratic, since computation of the normalizing constants $\sum_{\ell=1}^N \omega_s^\ell q(\xi_s^\ell, \xi_{s+1}^i)$ is required for all $i \in [\![1, N]\!]$ and $s \in [\![0, t-1]\!]$. Consequently, FFBSm is a computationally intensive approach.

2.3.3. *Forward-only implementation of FFBSm*

Appealingly, as noted by [9], in the case of additive state functionals the sequence $\{\phi_t^N h_t\}_{t \in \mathbb{N}}$ can be computed *on-the-fly* as t increases on the basis of the recursion (2.6). More specifically, plugging the estimates (2.8) into the recursion in question yields particle approximations $\{\tilde{\tau}_t^i\}_{i=1}^N$ of the statistics $\{\mathbf{T}_t h_t(\xi_t^i)\}_{i=1}^N$ evaluated at the particle locations. After initializing $\tilde{\tau}_0^i = 0$ for all $i \in [\![1, N]\!]$, these approximations may, when new observations become available, be updated by first evolving the particle filter sample one step and then setting

$$\tilde{\tau}_{t+1}^i = \sum_{j=1}^N \frac{\omega_t^j q(\xi_t^j, \xi_{t+1}^i)}{\sum_{\ell=1}^N \omega_t^\ell q(\xi_t^\ell, \xi_{t+1}^i)} \{\tilde{\tau}_t^j + \tilde{h}_t(\xi_t^j, \xi_{t+1}^i)\} \qquad (t \in \mathbb{N}), \tag{2.10}$$

yielding

$$\phi_{0:t|t}^N h_t = \sum_{i=1}^N \frac{\omega_t^i}{\Omega_t} \tilde{\tau}_t^i \tag{2.11}$$

as an estimate of $\phi_{0:t|t} h_t$. Note that (2.11) provides a particle interpretation of the backward decomposition (2.5). Besides allowing for online processing of the data, the algorithm has also the appealing property that only the current statistics $\{\tilde{\tau}_t^i\}_{i=1}^N$ and particle sample $\{(\xi_t^i, \omega_t^i)\}_{i=1}^N$ need to be stored in the memory. Still, the complexity of the scheme is $\mathcal{O}(N^2)$ due to the computation of the normalizing constants of the backward kernel induced by the particle filter.

2.3.4. *Forward-filtering backward-simulation* (*FFBSi*)

In order to remedy the high computational complexity of FFBSm, FFBSi generates trajectories on the *index space* $[\![1, N]\!]^{t+1}$ by simulating repeatedly a time-reversed, inhomogeneous Markov chain $\{\tilde{J}_s\}_{s=0}^t$ with transition probabilities

$$\mathbf{\Lambda}_s^N(i, j) := \frac{\omega_s^j q(\xi_s^j, \xi_{s+1}^i)}{\sum_{\ell=1}^N \omega_t^\ell q(\xi_s^\ell, \xi_{s+1}^i)} \qquad \big((s, i, j) \in [\![0, t-1]\!] \times [\![1, N]\!]^2\big) \tag{2.12}$$

and initial distribution (i.e., distribution at time t) $\mathsf{Pr}(\{\omega_t^j\}_{j=1}^N)$. Given $\{\tilde{J}_s\}_{s=0}^t$, an approximate draw from the joint smoothing distribution is formed by the random vector $(\xi_0^{\tilde{J}_0}, \dots, \xi_t^{\tilde{J}_t})$. Consequently, the uniformly weighted occupation measure associated with a set of conditionally independent such draws provides a finite-dimensional approximation of the smoothing distribution $\phi_{0:t|t}$; see [20]. In this basic formulation of FFBSi, the backward sampling pass requires the normalizing constants of the particle-based backward kernels to be computed, and hence the algorithm suffers from a quadratic complexity. On the other hand, on the contrary to FFBSm, this complexity is the same for *all* types of objective functions (whereas FFBSm has quadratic complexity only when applied to additive state functionals). However, following [11] it is, under the assumption that there exists $\bar{\varepsilon} \in \mathbb{R}_+^*$ such that $q(x, x') \leq \bar{\varepsilon}$ for all $(x, x') \in \mathsf{X}^2$ (an assumption that is satisfied for most models of interest), possible to reduce the computational complexity of FFBSi by simulating the approximate backward kernel using the following accept–reject technique. In order to sample from $\mathsf{Pr}(\{\mathbf{\Lambda}_s^N(i, j)\}_{j=1}^N)$ for given $s \in [\![0, t-1]\!]$ and $i \in [\![1, N]\!]$, a candidate J^* drawn from the proposal distribution $\mathsf{Pr}(\{\omega_s^j\}_{j=1}^N)$ is accepted with probability $q(\xi_s^{J^*}, \xi_{s+1}^i)/\bar{\varepsilon}$. The procedure is repeated until acceptance; see Algorithm 3 for an efficient way of implementing this approach. Under the additional assumption that the transition density is bounded also from below (see Assumption 2 below) it can be shown (see [11], Proposition 2) that the computational complexity of this accept–reject-based FFBSi algorithm is indeed *linear* (i.e., $\mathcal{O}(N)$).

3. Main results

Requiring separate forward and backward processing of the data, the standard design of FFBSi is not useful in online applications. We hence propose a novel algorithm which can be viewed as a hybrid between the forward-only implementation of the FFBSm algorithm and the FFBSi algorithm. In order to gain computational effort, it then replaces, in the spirit of FFBSi, exact computation of (2.10) by a Monte Carlo estimate. The algorithm, which is presented in the next section, is furnished with rigorous theoretical results concerning its convergence and numerical stability in Section 3.2.

3.1. The particle-based, rapid incremental smoother (PaRIS)

Given estimates $\{\tau_t^i\}_{i=1}^N$ of the auxiliary statistics $\{\mathbf{T}_t h_t(\xi_t^i)\}_{i=1}^N$ and a particle sample $\{(\xi_t^i, \omega_t^i)\}_{i=1}^N$ targeting the filter ϕ_t, the algorithm updates the estimated auxiliary statistics by, first,

Algorithm 2 Particle-based, rapid incremental smoother (PaRIS)

Require: Particles sample $\{(\xi_t^i, \omega_t^i)\}_{i=1}^N$ targeting ϕ_t and estimated auxiliary statistics $\{\tau_t^i\}_{i=1}^N$.

1: run $\{(\xi_{t+1}^i, \omega_{t+1}^i)\}_{i=1}^N \leftarrow \mathsf{PF}(\{(\xi_t^i, \omega_t^i)\}_{i=1}^N)$;
2: **for** $i = 1 \to N$ **do**
3:     **for** $j = 1 \to \tilde{N}$ **do**
4:         draw $J_{t+1}^{(i,j)} \sim \mathsf{Pr}(\{\omega_t^\ell q(\xi_t^\ell, \xi_{t+1}^i)\}_{\ell=1}^N)$;
5:     **end for**
6:     Set $\tau_{t+1}^i \leftarrow \tilde{N}^{-1} \sum_{j=1}^{\tilde{N}} (\tau_t^{J_{t+1}^{(i,j)}} + \tilde{h}_t(\xi_t^{J_{t+1}^{(i,j)}}, \xi_{t+1}^i))$;
7: **end for**
8: **return** $\{\tau_{t+1}^i\}_{i=1}^N$ and $\{(\xi_{t+1}^i, \omega_{t+1}^i)\}_{i=1}^N$.

propagating the particle cloud one step, yielding $\{(\xi_{t+1}^i, \omega_{t+1}^i)\}_{i=1}^N$, second, drawing, for each $i \in [\![1, N]\!]$, conditionally independent and identically distributed indices $\{J_{t+1}^{(i,j)}\}_{j=1}^{\tilde{N}}$, where $\tilde{N} \in \mathbb{N}^*$ is some given sample size referred to as the *precision parameter*, according to

$$\{J_{t+1}^{(i,j)}\}_{j=1}^{\tilde{N}} \sim \mathsf{Pr}\big(\{\mathbf{\Lambda}_t^N(i,\ell)\}_{\ell=1}^N\big)^{\otimes \tilde{N}} \qquad \big(i \in [\![1, N]\!]\big),$$

where the transition probabilities $\mathbf{\Lambda}_t^N$ are defined in (2.12), and third, letting

$$\tau_{t+1}^i = \tilde{N}^{-1} \sum_{j=1}^{\tilde{N}} \big(\tau_t^{J_{t+1}^{(i,j)}} + \tilde{h}_t\big(\xi_t^{J_{t+1}^{(i,j)}}, \xi_{t+1}^i\big)\big) \qquad \big(i \in [\![1, N]\!]\big).$$

Using the updated statistics $\{\tau_{t+1}^i\}_{i=1}^N$, an estimate of $\phi_{0:t+1|t+1} h_{t+1} = \phi_{t+1} \mathbf{T}_{t+1} h_{t+1}$ is obtained as $\sum_{i=1}^N \omega_{t+1}^i \tau_{t+1}^i / \Omega_{t+1}$. As for FFBSm, the algorithm is initialized by setting $\tau_0^i = 0$ for $i \in [\![1, N]\!]$. The resulting smoother, which is summarized in Algorithm 2, allows for online processing with constant memory requirements, as it requires only the current particle cloud and estimated auxiliary statistics to be stored at each iteration. In addition, applying, in Step (4), the accept–reject technique described in the previous section yields, for a given $\tilde{N}$, an algorithm with *linear complexity*.

In the PaRIS scheme, the precision parameter $\tilde{N}$ has to be set by the user. As shown in Section 3.2, the algorithm is asymptotically consistent (as the particle sample size N tends to infinity) *for any fixed* $\tilde{N} \in \mathbb{N}^*$ (i.e., the precision parameter does not need to be increased with N in order to guarantee consistency). Increasing the precision parameter increases the accuracy of the algorithm at the cost of additional computational complexity. Importantly, there is a *significant qualitative difference between the cases* $\tilde{N} = 1$ *and* $\tilde{N} \geq 2$, and it turns out that the latter is required to keep PaRIS numerically stable. This will be clear from the theoretical bounds on the asymptotic variance obtained in Section 3.2 as well as from the numerical experiments in Section 4.

In order to understand the fundamental difference between the cases $\tilde{N} = 1$ and $\tilde{N} \geq 2$, we may use the backward indices to connect the particles of different generations. Hence, let, for all

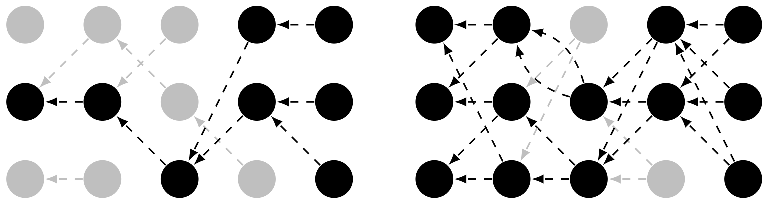

Figure 1. Genealogical traces corresponding to backward simulation in the PaRIS algorithm. Columns of nodes refer to different particle populations (with $N = 3$) at different time points (with time increasing rightward) and arrows indicate connections through the relation $\dashleftarrow$. Black-colored particles are included in the support S_4 of the final estimator, while gray-colored ones are inactive.

$t \in \mathbb{N}$ and $i \in [\![1, N]\!]$, $\mathcal{J}_{t,t}^{(i,\varnothing)} := i$ and, for all $s \in [\![0, t-1]\!]$ and $j_{s:t-1} \in [\![1, \tilde{N}]\!]^{t-s}$,

$$\mathcal{J}_{s,t}^{(i,j_{s:t-1})} := J_{s+1}^{(\mathcal{J}_{s+1,t}^{(i,j_{s+1:t-1})}, j_s)}$$

and let us write "$\xi_s^\ell \dashleftarrow \xi_t^i$" if there exists a sequence $j_{s:t-1}$ of indices such that $\ell = \mathcal{J}_{s,t}^{(i,j_{s:t-1})}$. Note that the support of the PaRIS estimator at time t is given by $\mathsf{S}_t := \prod_{s=0}^t \mathsf{A}_{s,t} \subset \mathsf{X}^{t+1}$, where $\mathsf{A}_{s,t} := \bigcup_{i=1}^N \{\xi_s^\ell : \xi_s^\ell \dashleftarrow \xi_t^i\} \subset \{\xi_s^\ell\}_{\ell=1}^N$ (so that, since $\xi_t^i \dashleftarrow \xi_t^i$ for all $i \in [\![1, N]\!]$, $\mathsf{A}_{t,t} = \{\xi_t^\ell\}_{\ell=1}^N$). When $\tilde{N} = 1$, the sequence $\{\#\mathsf{A}_{s,t}\}_{t=s}^\infty$ is non-decreasing, and $\#\mathsf{A}_{s,t} = 1 \Rightarrow \#\mathsf{A}_{u,t} = 1$ for all $u \in [\![0, s]\!]$. This implies a degeneracy phenomenon that resembles closely that of the poor man's smoother. On the contrary, in the case $\tilde{N} \geq 2$ it may well occur that $\#\mathsf{A}_{s,t} > \#\mathsf{A}_{s,t+1}$, also when $\#\mathsf{A}_{s,t+1} = 1$. The previous is, for $N = 3$ and $t = 4$, illustrated graphically in Figure 1, where columns of nodes represent particle clouds at different time steps (with time increasing rightward) and arrows indicate connections through the relation $\dashleftarrow$. Black-colored particles are included in the support S_4 of the final estimator, while gray-colored ones are inactive. As clear from Figure 1(a), setting $\tilde{N} = 1$ depletes quickly the support of the estimator, leading to a numerically unstable algorithm. Figure 1(b) shows the same configuration as in (a), but with one additional backward sample (i.e., $\tilde{N} = 2$). In this case, the sequence $\{\#\mathsf{A}_{s,4}\}_{s=0}^4$ is no longer non-decreasing, and a high degree of depletion at some time points (such as $s = 2$) has merely local effect of the support of the estimator. In the coming sections, the fact that PaRIS stays numerically stable for any fixed $\tilde{N} \geq 2$ is established theoretically as well as through simulations.

3.2. Theoretical results

The coming convergence analysis is driven by the following assumption.

Assumption 1.

(i) *For all $t \in \mathbb{N}$, the measure density $g_t \in \mathsf{F}_\mathrm{b}(\mathcal{X})$ is a positive bounded measurable function.*
(ii) *The transition density $q \in \mathsf{F}_\mathrm{b}(\mathcal{X}^2)$ is a bounded measurable function.*

Assumption 1(i) implies finiteness and positiveness of the particle weights; the boundedness of the transition density q implied by Assumption 1(ii) allows, besides certain technical arguments (formalized in Lemma 14) based on the generalized Lebesgue theorem, the accept–reject sampling technique discussed in Section 2.3.4 to be used.

It turns out to be necessary to establish the convergence of PaRIS for a slightly more general *affine modification* of the additive state functional (1.2) under consideration. More specifically, we will verify that for all $t \in \mathbb{N}$ and bounded measurable functions $(f_t, \tilde{f}_t) \in \mathsf{F}_b(\mathcal{X})^2$,

$$\sum_{i=1}^{N} \frac{\omega_t^i}{\Omega_t} \{ \tau_t^i f_t(\xi_t^i) + \tilde{f}_t(\xi_t^i) \} \overset{N\to\infty}{\simeq} \phi_t(\mathbf{T}_t h_t f_t + \tilde{f}_t), \tag{3.1}$$

where $\{\tau_t^i\}_{i=1}^N$ and $\{(\xi_t^i, \omega_t^i)\}_{i=1}^N$ are the output of Algorithm 2, in the senses of exponential concentration, weak convergence, and L_p error. The analogous results for the original additive state functional are then obtained as corollaries by simply applying (3.1) with $f_t \equiv \mathbb{1}_{\mathsf{X}}$ and $\tilde{f}_t \equiv \mathbb{1}_{\mathsf{X}^c}$. Our proofs, which are presented in Appendix A, are based on single-step analyses of the scheme and rely on techniques developed in [11] and [12]. Nevertheless, the analysis of PaRIS is, especially in the case of weak convergence, highly non-trivial due to the complex dependence between the ancestral lineages of the particles induced by the backward sampling approach (on the contrary to standard FFBSi, where the backward trajectories are conditionally independent; see the previous section).

3.2.1. *Hoeffding-type inequalities*

Besides being a result of independent interest, the following exponential concentration inequality for finite sample sizes N plays an instrumental role in the proof of the CLT in the next section. For reasons that will be clear in the proof of Theorem 3, the bound is established for the unnormalized as well as the normalized estimator; see Theorem 1, equations (i) and (ii), respectively.

Theorem 1. *Let Assumption 1 hold. Then for all $t \in \mathbb{N}$, bounded measurable functions $(f_t, \tilde{f}_t) \in \mathsf{F}_b(\mathcal{X})^2$, and $\tilde{N} \in \mathbb{N}^*$ there exist constants $(c_t, \tilde{c}_t) \in (\mathbb{R}_+^*)^2$ (depending on h_t, $\tilde{N}$, f_t, and $\tilde{f}_t$) such that for all $N \in \mathbb{N}^*$ and all $\varepsilon \in \mathbb{R}_+^*$,*

(i) $\mathbb{P}(|\frac{1}{N}\sum_{i=1}^N \omega_t^i \{\tau_t^i f_t(\xi_t^i) + \tilde{f}_t(\xi_t^i)\} - \phi_{t-1}\mathbf{L}_{t-1}(\mathbf{T}_t h_t f_t + \tilde{f}_t)| \geq \varepsilon) \leq c_t \exp(-\tilde{c}_t N \varepsilon^2)$,

(ii) $\mathbb{P}(|\sum_{i=1}^N \frac{\omega_t^i}{\Omega_t} \{\tau_t^i f_t(\xi_t^i) + \tilde{f}_t(\xi_t^i) - \phi_t(\mathbf{T}_t h_t f_t + \tilde{f}_t)\}| \geq \varepsilon) \leq c_t \exp(-\tilde{c}_t N \varepsilon^2)$

(*with the convention $\phi_{-1} \equiv \chi$*).

The following is an immediate consequence of Theorem 1.

Corollary 2. *Let Assumption 1 hold. Then for all $t \in \mathbb{N}$ and $\tilde{N} \in \mathbb{N}^*$ there exist constants $(c_t, \tilde{c}_t) \in (\mathbb{R}_+^*)^2$ (depending on h_t and $\tilde{N}$) such that for all $N \in \mathbb{N}^*$ and all $\varepsilon \in \mathbb{R}_+^*$,*

$$\mathbb{P}\left(\left| \sum_{i=1}^N \frac{\omega_t^i}{\Omega_t} \tau_t^i - \phi_{0:t|t} h_t \right| \geq \varepsilon \right) \leq c_t \exp\left(-\tilde{c}_t N \varepsilon^2\right).$$

3.2.2. *Central limit theorems and asymptotic L_p error*

Theorem 3. *Let Assumption 1 hold. Then for all $t \in \mathbb{N}$, bounded measurable functions $(f_t, \tilde{f}_t) \in \mathsf{F}_b(\mathcal{X})^2$, and $\tilde{N} \in \mathbb{N}^*$, as $N \to \infty$,*

$$\sqrt{N} \sum_{i=1}^{N} \frac{\omega_t^i}{\Omega_t} \{\tau_t^i f_t(\xi_t^i) + \tilde{f}_t(\xi_t^i) - \phi_t(\mathbf{T}_t h_t f_t + \tilde{f}_t)\} \xrightarrow{\mathcal{D}} \sigma_t \langle f_t, \tilde{f}_t \rangle (h) Z,$$

where Z has standard Gaussian distribution and

$$\begin{aligned} &\sigma_t^2 \langle f_t, \tilde{f}_t \rangle (h_t) \\ &\quad := \tilde{\sigma}_t^2 \langle f_t, \tilde{f}_t \rangle (h_t) \\ &\qquad + \sum_{s=0}^{t-1} \sum_{\ell=0}^{s} \tilde{N}^{\ell-(s+1)} \\ &\qquad \times \frac{\phi_\ell \mathbf{L}_\ell \{ \overleftarrow{\mathbf{Q}}_{\phi_\ell} (\mathbf{T}_\ell h_\ell + \tilde{h}_\ell - \mathbf{T}_{\ell+1} h_{\ell+1})^2 \mathbf{L}_{\ell+1} \cdots \mathbf{L}_s (g_{s+1} \{ \mathbf{L}_{s+1} \cdots \mathbf{L}_{t-1} f_t \}^2) \}}{(\phi_\ell \mathbf{L}_\ell \cdots \mathbf{L}_{s-1} \mathbb{1}_\mathsf{X})(\phi_s \mathbf{L}_s \cdots \mathbf{L}_{t-1} \mathbb{1}_\mathsf{X})^2} \end{aligned} \tag{3.2}$$

with

$$\tilde{\sigma}_t^2 \langle f_t, \tilde{f}_t \rangle (h_t) := \sum_{s=0}^{t-1} \frac{\phi_s \mathbf{L}_s \{ g_{s+1} \tilde{\mathbf{D}}_{s+1,t}^2 (h_t f_t + \tilde{f}_t) \}}{(\phi_s \mathbf{L}_s \cdots \mathbf{L}_{t-1} \mathbb{1}_\mathsf{X})^2}$$

being the asymptotic variance of the FFBSm algorithm (where, by convention, $\mathbf{L}_m \mathbf{L}_n = \mathrm{id}$ if $m > n$).

Remark 4. Since for all $s \in [\![0, t-1]\!]$ and $\ell \in [\![0, s]\!]$, $\ell - (s+1) \le -1$, it holds, in (3.2), that

$$\lim_{\tilde{N} \to \infty} \sigma_t^2 \langle f_t, \tilde{f}_t \rangle (h_t) = \tilde{\sigma}_t^2 \langle f_t, \tilde{f}_t \rangle (h_t),$$

that is, for large $\tilde{N}$ the asymptotic variance of PaRIS tends to that of the FFBSm algorithm. This is in line with our expectations, as the forward-only version of FFBSm can be viewed as a Rao–Blackwellization of PaRIS.

Again, the following is an immediate consequence of Theorem 3.

Corollary 5. *Let Assumption 1 hold. Then for all $t \in \mathbb{N}$ and $\tilde{N} \in \mathbb{N}^*$, as $N \to \infty$,*

$$\sqrt{N} \left(\sum_{i=1}^{N} \frac{\omega_t^i}{\Omega_t} \tau_t^i - \phi_{0:t|t} h_t \right) \xrightarrow{\mathcal{D}} \sigma_t (h_t) Z,$$

where Z has standard Gaussian distribution and

$$\begin{aligned}\sigma_t^2(h_t) &:= \tilde{\sigma}_t^2(h_t) \\ &\quad + \sum_{s=0}^{t-1} \sum_{\ell=0}^{s} \tilde{N}^{\ell-(s+1)} \\ &\quad \times \frac{\phi_\ell \mathbf{L}_\ell \{ \overleftarrow{\mathbf{Q}}_{\phi_\ell} (\mathbf{T}_\ell h_\ell + \tilde{h}_\ell - \mathbf{T}_{\ell+1} h_{\ell+1})^2 \mathbf{L}_{\ell+1} \cdots \mathbf{L}_s (g_{s+1} \{ \mathbf{L}_{s+1} \cdots \mathbf{L}_{t-1} \mathbb{1}_{\mathsf{X}} \}^2) \}}{(\phi_\ell \mathbf{L}_\ell \cdots \mathbf{L}_{s-1} \mathbb{1}_{\mathsf{X}})(\phi_s \mathbf{L}_s \cdots \mathbf{L}_{t-1} \mathbb{1}_{\mathsf{X}})^2}\end{aligned} \tag{3.3}$$

with

$$\tilde{\sigma}_t^2(h_t) := \sum_{s=0}^{t-1} \frac{\phi_s \mathbf{L}_s (g_{s+1} \tilde{\mathbf{D}}_{s+1,t}^2 h_t)}{(\phi_s \mathbf{L}_s \cdots \mathbf{L}_{t-1} \mathbb{1}_{\mathsf{X}})^2}$$

being the asymptotic variance of the FFBSm algorithm.

By following identically the lines of the proof of [13], Theorem 8, we may use Corollary 2 and Corollary 5 for deriving also the asymptotic L_p error of the estimates produced by the algorithm.

Corollary 6. *Let Assumption 1 hold. Then for all $p \in \mathbb{R}_+^*$, $t \in \mathbb{N}$, and $\tilde{N} \in \mathbb{N}^*$,*

$$\lim_{N \to \infty} \sqrt{N} \left\| \sum_{i=1}^{N} \frac{\omega_t^i}{\Omega_t} \tau_t^i - \phi_{0:t|t} h_t \right\|_{\mathsf{L}_p} = \sqrt{2} \sigma_t(h_t) \left(\frac{\Gamma\{(p+1)/2\}}{\sqrt{2\pi}} \right)^{1/p},$$

where $\sigma_t^2(h_t)$ is given in (3.3).

3.2.3. *Time uniform asymptotic variance bounds*

In the present section, we establish the long-term numerical stability of the PaRIS algorithm by bounding the asymptotic variance (3.3) (and hence, by Corollary 6, the asymptotic L_p error) using mixing-based arguments. We will treat separately joint smoothing and marginal smoothing, and derive, for precision parameters $\tilde{N} \geq 2$, $\mathcal{O}(t)$ and $\mathcal{O}(1)$ bounds, respectively, on the asymptotic variances in these cases. Since such time dependence is the best possible for SMC error bounds on the path and marginal spaces, these results confirm the conjecture that the algorithm stays numerically stable for precision parameters of this sort. Similar results for the FFBSm and FFBSi algorithms were obtained in [11,18]. The analysis will be carried through under the following *strong mixing* assumption, which is standard in the literature of SMC analysis (see [10] and, e.g., [5,7,8,13] for refinements) and points to applications where the state space X is a compact set.

Assumption 2.

(i) *There exist constants $0 < \underline{\varepsilon} < \bar{\varepsilon} < \infty$ such that for all $(x, \tilde{x}) \in \mathsf{X}^2$,*

$$\underline{\varepsilon} \leq q(x, \tilde{x}) \leq \bar{\varepsilon},$$

and we define $\varrho := 1 - \underline{\varepsilon}/\bar{\varepsilon}$.

(ii) *There exist constants* $0 < \underline{\delta} < \bar{\delta} < \infty$ *such that for all* $t \in \mathbb{N}$, $\|g_t\|_\infty \le \bar{\delta}$ *and* $\underline{\delta} \le \mathbf{L}_t \mathbb{1}_{\mathsf{X}}(x)$, $x \in \mathsf{X}$.

Joint smoothing. The following assumption implies that the additive functional under consideration grows at most linearly with time, which is a minimal requirement for obtaining an $\mathcal{O}(t)$ asymptotic variance.

Assumption 3. *There exists* $|\tilde{h}|_\infty \in \mathbb{R}_+^*$ *such that for all* $s \in \mathbb{N}$, $\operatorname{osc}(\tilde{h}_s) \le |\tilde{h}|_\infty$.

As an auxiliary result, we provide an $\mathcal{O}(t)$ bound on the asymptotic variance of the FFBSm algorithm; see [18] for a similar result on the L_p error for finite particle sample sizes.

Proposition 7. *Let Assumption 2 and Assumption 3 hold. Then*

$$\limsup_{t\to\infty} \frac{1}{t} \tilde{\sigma}_t^2(h_t) \le |\tilde{h}|_\infty^2 \frac{4\bar{\delta}}{\underline{\delta}(1-\varrho)^4}.$$

In the light of Proposition 7, it suffices to bound the second term of (3.3) by a quantity of order $\mathcal{O}(t)$. This yields the following result, where interestingly, the incremental asymptotic variance caused by the backward simulation is *inversely proportional to the precision parameter* $\tilde{N}$. This is well in line with the theory of *random weight SMC methods*, in which, in similarity to our algorithm, intractable quantities (the importance weights) are replaced by random and unbiased estimates of the same (see [31,32]).

Theorem 8. *Let Assumption 2 and Assumption 3 hold. Then for all* $\tilde{N} \ge 2$,

$$\limsup_{t\to\infty} \frac{1}{t} \sigma_t^2(h_t) \le |\tilde{h}|_\infty^2 \frac{\bar{\delta}}{\underline{\delta}(1-\varrho)^4} \left(4 + \frac{(4\bar{\delta}\varrho^2 + 1 - \varrho)^2}{(\tilde{N}-1)(1-\varrho)}\right),$$

where $\sigma_t^2(h_t)$ *is defined in* (3.3).

Marginal smoothing. We turn to marginal smoothing, that is, the situation when all terms of the additive functional are zero but a single one. For such a particular objective function, we are able to construct a time uniform bound on the *unnormalized* asymptotic variance of the same form as before, with one term representing the FFBSm asymptotic variance (see [11], Theorem 12) and one additional term being inversely proportional to the precision parameter and representing the loss of accuracy introduced by backward sampling.

Assumption 4. *The additive functional has the following form. For some* $\hat{s} \in \mathbb{N}$,

$$\tilde{h}_s(x_{s:s+1}) = \begin{cases} 0 & \text{for } s \ne \hat{s}, \\ \tilde{h}_{\hat{s}}(x_{\hat{s}}) & \text{for } s = \hat{s}. \end{cases}$$

Theorem 9. *Let Assumption 2 and Assumption 4 hold. Then for all $t \in \mathbb{N}$ and $\tilde{N} \geq 2$,*

$$\sigma_t^2(h_t) \leq \operatorname{osc}^2(\tilde{h}_{\hat{s}}) \frac{\bar{\delta}}{(1-\varrho)^3} \left(\bar{\delta} \frac{1+\varrho^2}{1+\varrho} + 4 \frac{1}{\underline{\delta}(\tilde{N}-1)\{(1-\varrho^2) \wedge (1/2)\}} \right),$$

where $\sigma_t^2(h_t)$ is defined in (3.3).

3.2.4. *Computational complexity*

We conclude this section with some comments on the complexity of the algorithm. Under Assumption 1(ii), we may cast the accept–reject technique proposed in [11], Algorithm 1, into the framework of PaRIS. A pseudo-code describing the resulting scheme is provided by Algorithm 3 in Section B.2. For a given $t \in \mathbb{N}$, we denote by $C_t\langle N, \tilde{N}\rangle$ the (random) number of elementary operations needed for executing the PaRIS algorithm parameterized by $(N, \tilde{N}) \in (\mathbb{N}^*)^2$ from time zero to time t. Note that $C_t\langle N, \tilde{N}\rangle$ is strongly data dependent, as the observations $\{y_s\}_{s=0}^t$ effect, via the particle weights, the acceptance probabilities at the different time steps. Still, under the strong mixing assumption above it is possible to bound uniformly this random variable. The following result is an immediate consequence of [11], Proposition 2.

Theorem 10. *Let Assumption 2 hold. Then there exists a constant $c \in \mathbb{R}_+^*$ such that $\mathbb{E}[C_t\langle N, \tilde{N}\rangle] \leq ctN\tilde{N}/(1-\varrho)$ for all $t \in \mathbb{N}$.*

Thus, the expected number of trials grows linearly with time, the number of particles, and the precision parameter, showing the importance of keeping the latter at a minimum. On the other hand, since the variance bound derived in Proposition 7 (and Theorem 9) is inversely proportional to $\tilde{N}$, using an excessively large precision parameter will not pay off in terms of variance reduction (as the variance term controlled by the precision parameter will be negligible beside the variance corresponding to FFBSm). *We hence advocate keeping $\tilde{N}$ at a moderate value*, and will return to this matter in connection to the numerical illustrations of the next section.

4. Simulations

An exhaustive study of the numerical aspects of PaRIS is beyond the scope of the present paper; nevertheless, we benchmark the algorithm on two different models, namely

- a linear Gaussian state-space model (for which all quantities of interest can be computed exactly for comparison) and
- a stochastic volatility model [22].

4.1. Linear Gaussian state-space model

We first consider the linear Gaussian state-space model

$$\begin{aligned} X_{t+1} &= aX_t + \sigma_\varepsilon \varepsilon_{t+1}, \\ Y_t &= bX_t + \sigma_\zeta \zeta_t \qquad (t \in \mathbb{N}), \end{aligned} \tag{4.1}$$

where $\mathsf{Y} = \mathsf{X} = \mathbb{R}$ and $\{\varepsilon_t\}_{t\in\mathbb{N}^*}$ and $\{\zeta_t\}_{t\in\mathbb{N}}$ are sequences of mutually independent standard normally distributed random variables. The parameters $(a, b) \in \mathbb{R}^2$ and $(\sigma_\varepsilon, \sigma_\zeta) \in (\mathbb{R}_+^*)^2$ are considered to be known. We aim at computing smoothed expectations of the sufficient statistics

$$\begin{aligned} h_t^{(1)}(x_{0:t}) &:= \sum_{s=0}^{t} x_s, \qquad h_t^{(2)}(x_{0:t}) := \sum_{s=0}^{t} x_s^2, \\ h_t^{(3)}(x_{0:t}) &= \sum_{s=0}^{t-1} x_s x_{s+1} \qquad \left(x_{0:t} \in \mathsf{X}^{t+1}\right) \end{aligned} \tag{4.2}$$

under the dynamics governed by the parameter vector $(a, b, \sigma_\varepsilon, \sigma_\zeta) = (0.7, 1, 0.2, 1)$, and assume for simplicity that the model is well-specified. For this model, the *disturbance smoother* (see, e.g., [5], Algorithm 5.2.15) provides the exact values of the smoothed sufficient statistics, and we compared these values with approximations obtained using PaRIS as well as the forward-only implementation of FFBSm. With our implementation, parameterizing PaRIS and FFBSm with $(N, \tilde{N}) = (150, 2)$ and $N = 50$, respectively, resulted in very similar computational times for the two algorithms, with PaRIS being slightly faster (recall that FFBSm has a quadratic complexity). As clear from the box plots (based on *time-normalized* estimates) displayed in Figure 2, PaRIS outperforms clearly FFBSm as the former exhibits lower variance as well as smaller bias for equal computational time.

As a measure of numerical performance, we define *efficiency* as inverse sample variance over computational time. Figure 3 reports the efficiencies by which the PaRIS and forward-only FF-BSm algorithms estimate $\phi_{0:t|t} h_t^{(1)}$ using each $N = 500$ particles. As evident from the plot, PaRIS exhibits a higher efficiency uniformly over all time points. The variance estimates were based on 50 replicates.

In order to examine the dependence of the performance of PaRIS on the design of the precision parameter $\tilde{N}$, we produced estimates of $\phi_{0:t|t} h_t^{(1)}$ for $t \in [\![0, 1000]\!]$ using the algorithm for each of the precision parameters $\tilde{N} \in \{1, 2, 3, 4, 10, 30\}$. All these estimators were computed on the basis of the same forward particles, so also an additional FFBSm-based estimator. This experiment was, in order to estimate the variances of the (seven) different estimators, replicated 100 times for the same fixed sequence of observations. Figure 4, displaying estimated variance as a function of time, shows a momentous difference between the cases $\tilde{N} = 1$ and $\tilde{N} > 1$ (note the difference in y-axis scale between the two graphs); the graphs in the top ($\tilde{N} = 1$) and bottom ($\tilde{N} > 1$) figures exhibit variance growths that appear to be close to quadratic and linear, respectively, which is well in accordance with the theory. Increasing the precision parameter $\tilde{N}$ from 2 to 4 implies some decrease of variance, while increasing the same from 4 to 30 has only marginal effect on the accuracy of the estimator (the difference between the variances corresponding to $\tilde{N} = 10$ and $\tilde{N} = 30$ is close to indistinguishable). This is perfectly in line with the theoretical results obtained in Section 3, where the second term of the variance bound in Theorem 8 is inversely proportional to the precision parameter. Finally, ratios of variances of estimators associated with different $\tilde{N}$ are displayed in Figure 5, which shows a linearly increasing ratio of the variances associated with $\tilde{N} = 1$ and $\tilde{N} = 2$ and a close to constant ratio of the variances associated with $\tilde{N} = 2$ and $\tilde{N} = 3$.

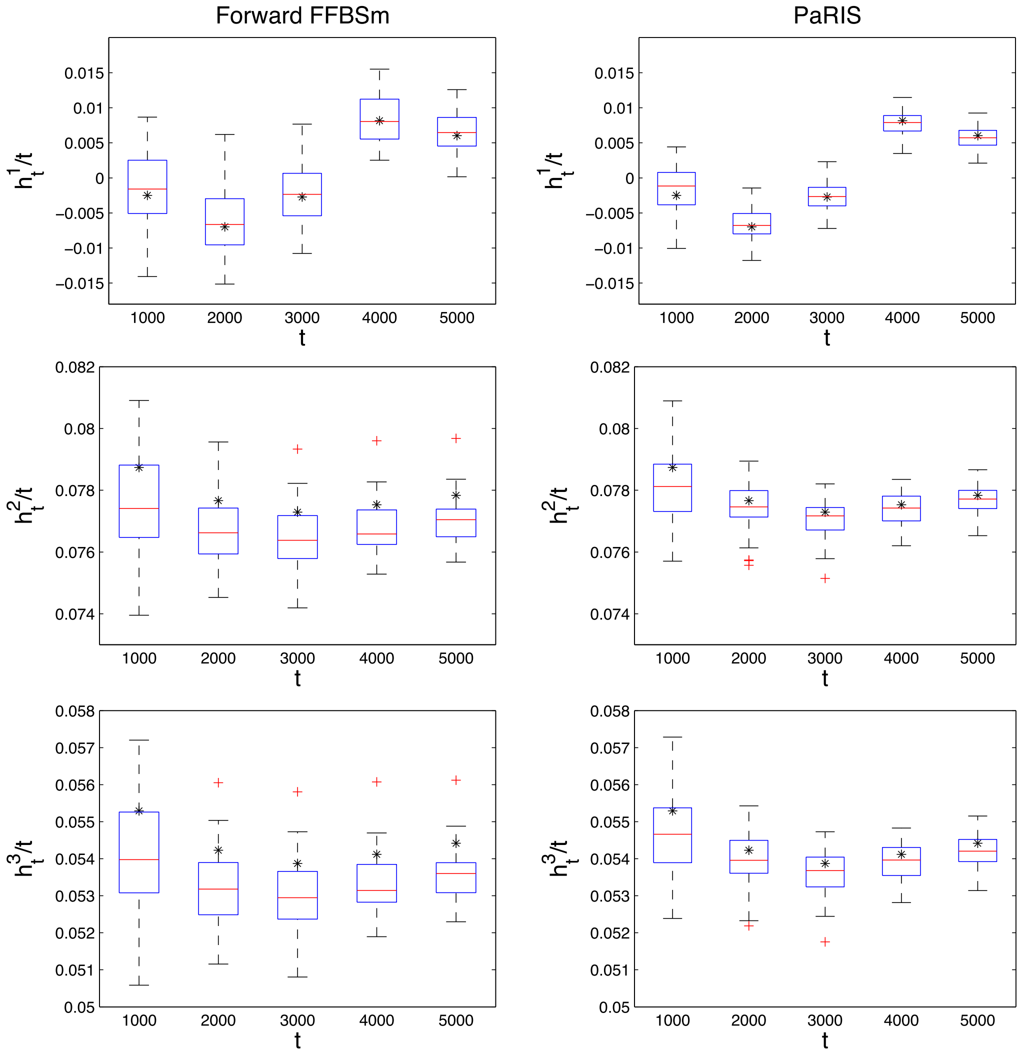

Figure 2. Box plots of estimates of smoothed sufficient statistics (4.2) for the linear Gaussian model (4.1) produced by PaRIS (right column) and the forward-only version of FFBSm (left column) using $(N, \tilde{N}) = (150, 2)$ and $N = 50$, respectively (yielding close to identical computational times). The boxes are based on 50 replicates of the estimates for the same fixed observation sequence and asterisks indicate exact values obtained with the disturbance smoother.

Finally, in order to illustrate our algorithm's capacity of coping with particle path degeneracy, we report, in Figure 6, the ratios $\#\mathsf{S}_t/\{N(t+1)\}$, $t \in [\![0, 1000]\!]$, where $\#\mathsf{S}_t$ is the cardinality of the support of the PaRIS algorithm at time t (in the notation of Section 3.1), for the precision parameters $\tilde{N} \in \{1, 2, 3, 10, 30\}$. Here, $N = 100$, and again the estimators associated with different

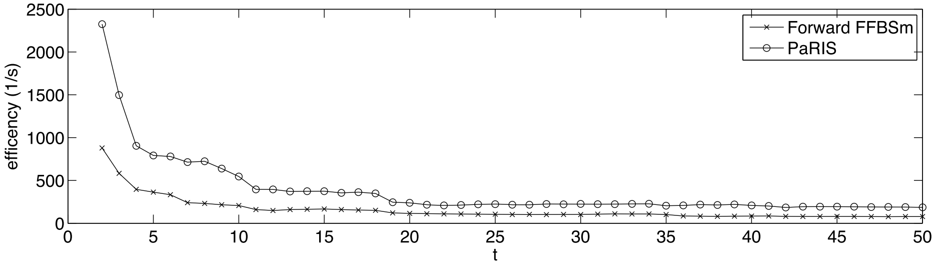

Figure 3. Estimated efficiencies for the PaRIS and forward-only FFBSm algorithms using each $N = 500$ particles. (The first time step is removed from the plot due to very high efficiencies for both algorithms.)

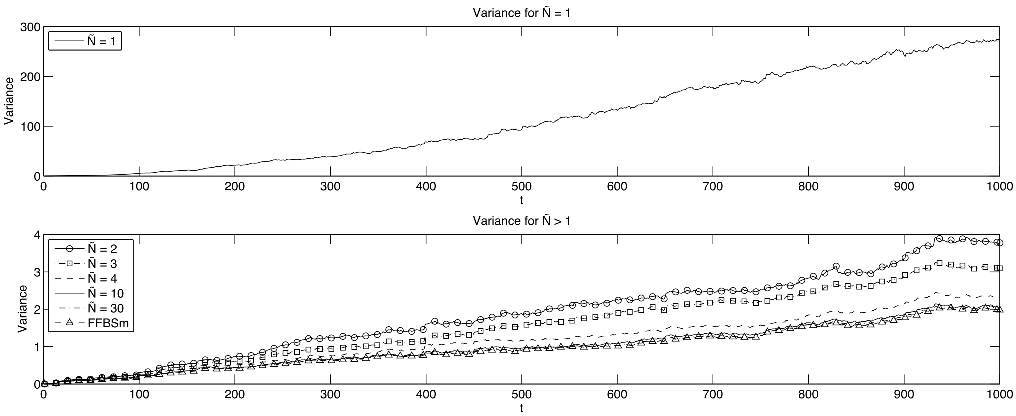

Figure 4. Estimated variances of PaRIS estimators for $\tilde{N} = 1$ (top graph) and $\tilde{N} \in \{2, 3, 4, 10, 30\}$ (bottom graph) at different time steps $t \in [\![0, 1000]\!]$. The bottom graph includes variance estimates of the forward-only FFBSm estimator. The variance estimates are based on 100 replicates.

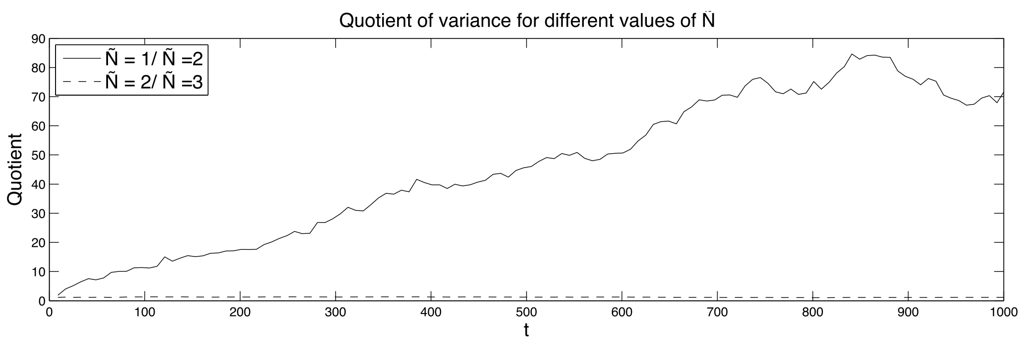

Figure 5. Ratios of variances of the estimators associated with $\tilde{N} = 1$ and $\tilde{N} = 2$ (solid line) and $\tilde{N} = 2$ and $\tilde{N} = 3$ (dashed line).

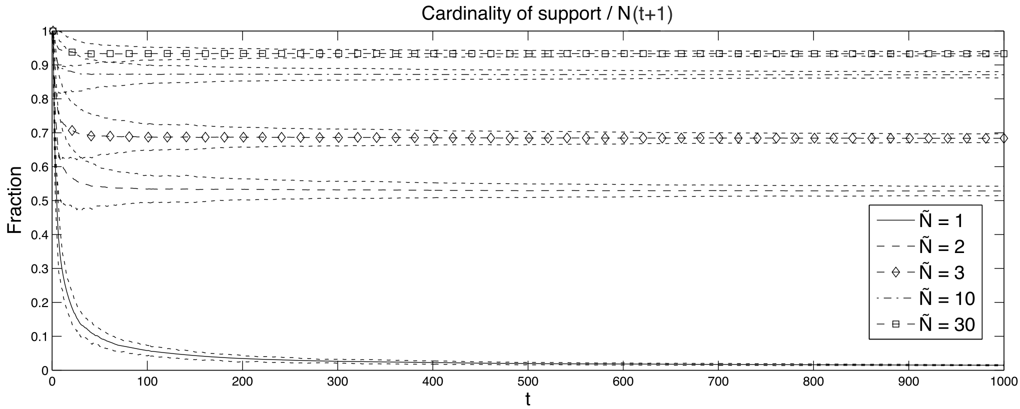

Figure 6. Plot the ratios $\#\mathsf{S}_t/\{N(t+1)\}$, $t \in [\![0, 1000]\!]$, for the precision parameters $\tilde{N} \in \{1, 2, 3, 10, 30\}$ and $N = 100$. The 95% confidence bounds are obtained on the basis of 100 replicates of the observation record.

precision parameters were based on the same forward particles. The 95% confidence bounds displayed the same plot were obtained on the basis of 100 replicates of observation record. Judging by these confidence bounds, the dependence of the copiousness of the support on the observations is fairly robust. Interestingly, for $\tilde{N} = 1$ the sequence of ratios tends quickly to zero, while letting $\tilde{N} > 1$ stabilizes completely the support of the estimator. Already $\tilde{N} = 2$ yields a support that involves, on the average and in the long run, more than 50% of all forward particles. Again, increasing the precision parameter has some effect for moderate values of the same, say, up to $\tilde{N} = 10$, while increasing the parameter further from 10 to 30 (which implies a significant increase of computational overhead) effects only marginally the cardinality of the support. Also this observation is perfectly in line with the theory presented in Section 3, consolidating our apprehension that only a modest value of $\tilde{N}$ is required as long as $\tilde{N} \geq 2$.

4.2. Stochastic volatility model

For the sake of completeness, we also consider a non-linear model, namely the standard stochastic volatility model

$$\begin{aligned} X_{t+1} &= \phi X_t + \sigma_\varepsilon \varepsilon_{t+1}, \\ Y_t &= \beta \exp(X_t/2)\zeta_t \qquad (t \in \mathbb{N}), \end{aligned} \tag{4.3}$$

where $\mathsf{X} = \mathsf{Y} = \mathbb{R}$ and $\{\varepsilon_t\}_{t\in\mathbb{N}^*}$ and $\{\zeta_t\}_{t\in\mathbb{N}}$ are as in the previous example. We assume that the model parameters $\phi \in \mathbb{R}$ and $(\sigma, \beta) \in (\mathbb{R}_+^*)^2$ are known and that the model is well-specified. Our aim is to compute, using again PaRIS and the forward-only implementation of FFBSm, smoothed

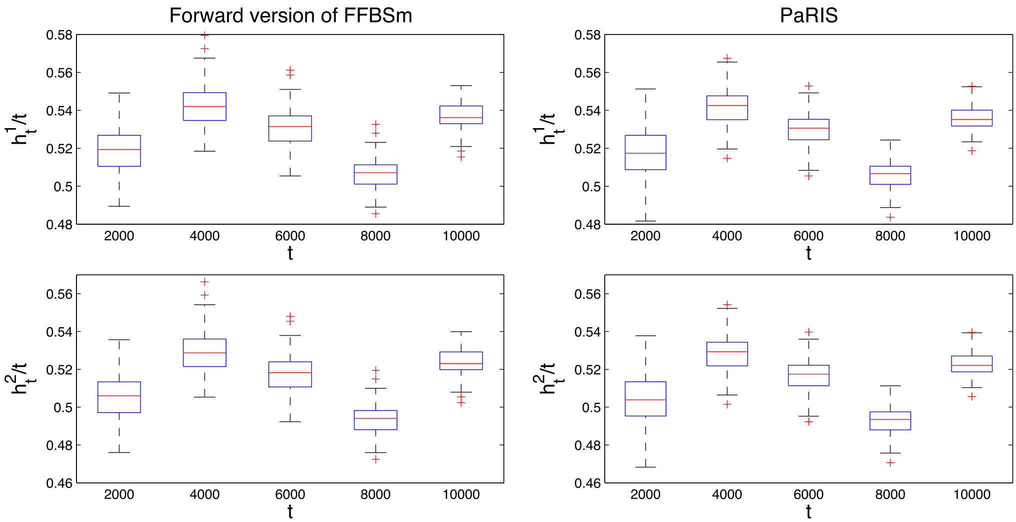

Figure 7. Box plots of estimates of smoothed sufficient statistics (4.4) for the stochastic volatility model (4.3) produced by PaRIS (right column) and the forward-only version of FFBSm (left column) using $(N, \tilde{N}) = (250, 2)$ and $N = 250$, respectively. With this parameterization, PaRIS was 5 times faster than the FFBSm algorithm. The boxes are based on 100 replicates of the estimates for the same fixed observation sequence.

expectations of the sufficient statistics

$$h_t^{(1)}(x_{0:t}) := \sum_{s=0}^{t} x_s^2, \qquad h_t^{(2)}(x_{0:t}) := \sum_{s=0}^{s-1} x_s x_{s+1} \qquad \left(x_{0:t} \in \mathsf{X}^{t+1}\right) \tag{4.4}$$

for a model parameterized by $(\phi, \sigma, \beta) = (0.975, 0.16, 0.63)$. In this case, both algorithms used $N = 250$ particles and the precision parameter of PaRIS was set to $\tilde{N} = 2$. Figure 7 shows box plots based on 100 replicates of estimates of $\phi_{0:t|t} h_t^{(i)}/t$, for $i \in \{1, 2\}$ and $t \in \{2000, 4000, 6000, 8000, 10\,000\}$, obtained using these methods. Even though the variance and the bias of the estimates produced by the two algorithms are comparable, PaRIS was now *5 times faster* than the FFBSm algorithm.

4.3. Some comments on the implementation

When applying accept–reject-based backward sampling (Algorithm 3), some acceptance probabilities will be small due to the random support of the particle-based backward kernel. In order to avoid getting stuck, it may be convenient to equip the algorithm with a threshold for the number of trials used at each accept–reject operation; when the threshold is reached, accept–reject sampling is cancelled and replaced by a draw from original distribution (recall that we are just using

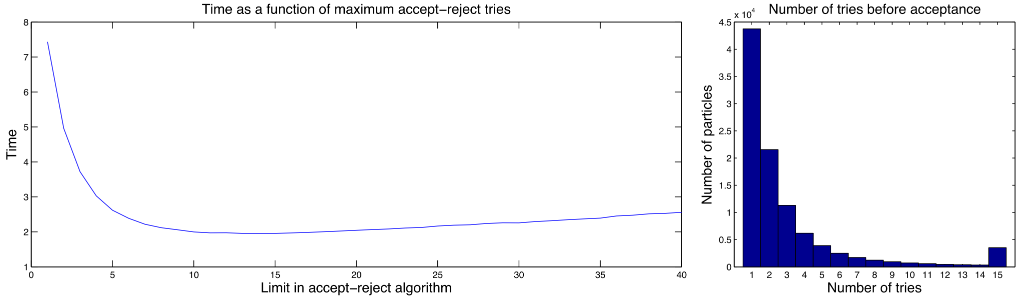

Figure 8. Computational time as a function of the size of the accept–reject threshold for the linear Gaussian model and $N = 250$ particles (left panel). The histogram to the right displays the number of trials needed before acceptance at any accept–reject sampling operation in algorithm when the threshold is 14 (corresponding to minimal computational time); the bar at 15 represents 3.55% of the occasions.

accept–reject sampling in order to reduce the computational work). Figure 8 displays computational time as a function of the size of this threshold for the linear Gaussian model and $N = 250$ particles. Interestingly, the graph has a minimum for the threshold value 14, and using this value we run the algorithm and counted the number of trials at any accept–reject sampling operation. The outcome is presented in the histogram plot to the right, from which it is clear that the majority of the particles are accepted after just a few trials (moreover, an index is most commonly accepted at once). In addition, at only 3.55% of the occasions, the number of trials exceeded the threshold. Needless to say, the optimal threshold depends on the model as well as the number of particles (when the number of particles is small, a too high threshold may have significant negative effect on the computational efficiency; on the contrary, when the number of particles is large, the performance of the algorithm is relatively robust vis-à-vis the design of the threshold). Further simulations not presented here indicate however that a threshold value around $\sqrt{N}$ could be a rule of thumb.

5. Conclusions

We have presented a novel algorithm, the particle-based, rapid incremental smoother, PaRIS, for computationally efficient online smoothing of additive state functionals in general HMMs. The algorithm, which is based on a backward decomposition of the smoothing distribution which can be implemented recursively for objective functions of additive type, can be viewed as a hybrid between the forward-only implementation of FFBSm and the FFBSi algorithm; more specifically, forward-only FFBSm may be viewed as a Rao–Blackwellized version of PaRIS. The algorithm is furnished with a number of convergence results, where the main result is a CLT of PaRIS's Monte Carlo output at the rate $\sqrt{N}$. The analysis of PaRIS is considerably more involved than that of the FFBSi algorithm due to the complex dependence structure introduced by the retrospective simulation (on the contrary to FFBSi, where the trajectories are conditionally independent given the particles generated in the forward pass). Interestingly, the design of the precision parameter, that is, the number of Monte Carlo simulations used for approximating the backward decomposition, turns out to be critical, since using a single backward draw yields a degeneracy phenomenon that resembles closely that of the poor man's smoother. However, as established theoretically as well as through simulations, using at least *two* such draws stabilizes completely the support of the estimator. For $\tilde{N} \geq 2$, we are able to derive $\mathcal{O}(1 + 1/(\tilde{N} - 1))$ and $\mathcal{O}(t\{1 + 1/(\tilde{N} - 1)\})$ bounds on the asymptotic variance in the cases of marginal and joint smoothing, respectively, and since the second term of these bounds is inversely proportional to the precision parameter, we suggest this parameter to be kept at a moderate value in order to gain computational speed. As known to the authors, this is the first analysis ever of this kind.

The algorithm we propose has a linear complexity in the number of particles while the forward-only implementation of FFBSm has a quadratic complexity, and a numerical comparison between the two shows clearly that PaRIS achieves the same accuracy as FFBSm at a considerably lower computational cost. In addition, similarly to forward-only FFBSm, our smoother has limited and constant memory requirements, as it needs only the current particle sample and a set of estimated auxiliary statistics to be stored at each iteration.

Smoothing of additive state functionals is a key ingredient of most – frequentistic or Bayesian – online parameter estimation techniques for HMMs. Since these applications are most often characterized by strict computational requirements, PaRIS can be naturally cast into any such framework.

Appendix A: Proofs

The following filtrations will be used repeatedly in the proofs. For all $(N, \tilde{N}) \in (\mathbb{N}^*)^2$, define

$$\tilde{\mathcal{F}}_t^N := \begin{cases} \sigma\left(\{\xi_0^i\}_{i=1}^N\right) & \text{for } t = 0, \\ \sigma\left(\xi_0^i, \xi_s^i, I_s^i, \{J_s^{(i,j)}\}_{j=1}^{\tilde{N}}; s \in [\![1, t]\!], i \in [\![1, N]\!]\right) & \text{for } t \in \mathbb{N}^*, \end{cases}$$

that is, $\tilde{\mathcal{F}}_t^N$ is the σ-field generated by all random variables produced during the first t iterations of PaRIS (Algorithm 2). (Since we consider exclusively the convergence of PaRIS as the particle sample size N tends to infinity for a *fixed* precision parameter $\tilde{N}$, we have omitted the latter from the notation.) In addition, for all $(N, \tilde{N}) \in (\mathbb{N}^*)^2$, let

$$\mathcal{F}_t^N := \begin{cases} \tilde{\mathcal{F}}_0^N & \text{for } t = 0, \\ \tilde{\mathcal{F}}_{t-1}^N \vee \sigma\left(\{\xi_t^i, I_t^i\}_{i=1}^N\right) & \text{for } t \in \mathbb{N}^*, \end{cases}$$

that is, $\mathcal{F}_t^N$ is the σ-field generated by all random variables produced by PaRIS up to Step (1) in the tth iteration. Note that for all $t \in \mathbb{N}$, $\tilde{\mathcal{F}}_t^N \subset \mathcal{F}_{t+1}^N$.

A.1. Two prefatory lemmas

Lemma 11. *For all $t \in \mathbb{N}$ and $(f_{t+1}, \tilde{f}_{t+1}) \in \mathsf{F}_\mathrm{b}(\mathcal{X})^2$ it holds that*

$$\phi_{t+1}(\mathbf{T}_{t+1} h_{t+1} f_{t+1} + \tilde{f}_{t+1}) = \frac{\phi_t\{\mathbf{T}_t h_t \mathbf{L}_t f_{t+1} + \mathbf{L}_t(\tilde{h}_t f_{t+1} + \tilde{f}_{t+1})\}}{\phi_t \mathbf{L}_t \mathbb{1}_\mathsf{X}}.$$

Proof. By combining the definitions of $\mathbf{T}_t$, $\mathbf{L}_t$ and using reversibility,

$$\begin{aligned}\phi_t \mathbf{L}_t(\mathbf{T}_{t+1} h_{t+1} f_{t+1}) &= \phi_t \mathbf{Q}\{\overleftarrow{\mathbf{Q}}_{\phi_t}(\mathbf{T}_t h_t + \tilde{h}_t) g_{t+1} f_{t+1}\} \\ &= \phi_t \mathbf{Q} \otimes \overleftarrow{\mathbf{Q}}_{\phi_t}\{(\mathbf{T}_t h_t + \tilde{h}_t) g_{t+1} f_{t+1}\} \\ &= \phi_t \otimes \mathbf{Q}\{(\mathbf{T}_t h_t + \tilde{h}_t) g_{t+1} f_{t+1}\} \\ &= \phi_t\{\mathbf{T}_t h_t \mathbf{L}_t f_{t+1} + \mathbf{L}_t(\tilde{h}_t f_{t+1})\}.\end{aligned}$$

Now the statement of the lemma follows by dividing both sides of the previous equation by $\phi_t \mathbf{L}_t \mathbb{1}_\mathsf{X}$ and using the identity $\phi_{t+1} = \phi_t \mathbf{L}_t / \phi_t \mathbf{L}_t \mathbb{1}_\mathsf{X}$. □

Lemma 12. *For all $t \in \mathbb{N}$, $(f_{t+1}, \tilde{f}_{t+1}) \in \mathsf{F}_\mathrm{b}(\mathcal{X})^2$, and $(N, \tilde{N}) \in (\mathbb{N}^*)^2$ the random variables $\{\omega_{t+1}^i(\tau_{t+1}^i f_{t+1}(\xi_{t+1}^i) + \tilde{f}_{t+1}(\xi_{t+1}^i))\}_{i=1}^N$ are, conditionally on $\tilde{\mathcal{F}}_t^N$, i.i.d. with expectation*

$$\begin{aligned}&\mathbb{E}[\omega_{t+1}^1\{\tau_{t+1}^1 f_{t+1}(\xi_{t+1}^1) + \tilde{f}_{t+1}(\xi_{t+1}^1)\} \mid \tilde{\mathcal{F}}_t^N] \\ &\quad = \sum_{i=1}^N \frac{\omega_t^i}{\Omega_t}\{\tau_t^i \mathbf{L}_t f_{t+1}(\xi_t^i) + \mathbf{L}_t(\tilde{h}_t f_{t+1} + \tilde{f}_{t+1})(\xi_t^i)\}.\end{aligned} \tag{A.1}$$

Proof. The multinomial selection procedure implies that the particles $\{\xi_{t+1}^i\}_{i=1}^N$ are i.i.d. conditionally on $\tilde{\mathcal{F}}_t^N$. Hence, since also the backward indices $\{J_{t+1}^{(i,j)}\}_{j=1}^{\tilde{N}}$ are i.i.d. conditionally on the particle ξ_t^i and the σ-field $\tilde{\mathcal{F}}_t^N$, we conclude that $\{\omega_{t+1}^i(\tau_{t+1}^i f_{t+1}(\xi_{t+1}^i) + \tilde{f}_{t+1}(\xi_{t+1}^i))\}_{i=1}^N$ are i.i.d. conditionally on $\tilde{\mathcal{F}}_t^N$.

In order to compute the common conditional expectation, we decompose the same according to

$$\begin{aligned}&\mathbb{E}[\omega_{t+1}^1\{\tau_{t+1}^1 f_{t+1}(\xi_{t+1}^1) + \tilde{f}_{t+1}(\xi_{t+1}^1)\} \mid \tilde{\mathcal{F}}_t^N] \\ &\quad = \mathbb{E}[\omega_{t+1}^1 \tau_{t+1}^1 f_{t+1}(\xi_{t+1}^1) \mid \tilde{\mathcal{F}}_t^N] + \mathbb{E}[\omega_{t+1}^1 \tilde{f}_{t+1}(\xi_{t+1}^1) \mid \tilde{\mathcal{F}}_t^N],\end{aligned}$$

where, by the tower property,

$$\begin{aligned}&\mathbb{E}[\omega_{t+1}^1 \tau_{t+1}^1 f_{t+1}(\xi_{t+1}^1) \mid \tilde{\mathcal{F}}_t^N] \\ &\quad = \mathbb{E}[\omega_{t+1}^1 f_{t+1}(\xi_{t+1}^1) \mathbb{E}[\tau_{t+1}^1 \mid \mathcal{F}_{t+1}^N] \mid \tilde{\mathcal{F}}_t^N] \\ &\quad = \mathbb{E}\left[\omega_{t+1}^1 f_{t+1}(\xi_{t+1}^1) \sum_{\ell=1}^N \frac{\omega_t^\ell q(\xi_t^\ell, \xi_{t+1}^1)}{\sum_{\ell'=1}^N \omega_t^{\ell'} q(\xi_t^{\ell'}, \xi_{t+1}^1)}\{\tau_t^\ell + \tilde{h}_t(\xi_t^\ell, \xi_{t+1}^1)\} \,\middle|\, \tilde{\mathcal{F}}_t^N\right] \\ &\quad = \sum_{i=1}^N \frac{\omega_t^i}{\Omega_t} \int q(\xi_t^i, x) g_{t+1}(x) f_{t+1}(x) \sum_{\ell=1}^N \frac{\omega_t^\ell q(\xi_t^\ell, x)}{\sum_{\ell'=1}^N \omega_t^{\ell'} q(\xi_t^{\ell'}, x)}\{\tau_t^\ell + \tilde{h}_t(\xi_t^\ell, x)\} \mu(\mathrm{d}x) \\ &\quad = \sum_{\ell=1}^N \frac{\omega_t^\ell}{\Omega_t}\{\tau_t^\ell \mathbf{L}_t f_{t+1}(\xi_t^\ell) + \mathbf{L}_t(\tilde{h}_t f_{t+1})(\xi_t^\ell)\}.\end{aligned}$$

We conclude the proof by noting that

$$\mathbb{E}\big[\omega_{t+1}^{1} \tilde{f}_{t+1}\big(\xi_{t+1}^{1}\big) \mid \tilde{\mathcal{F}}_t^N\big] = \sum_{\ell=1}^{N} \frac{\omega_t^\ell}{\Omega_t} \mathbf{Q}(g_{t+1} \tilde{f}_{t+1})\big(\xi_t^\ell\big) = \sum_{\ell=1}^{N} \frac{\omega_t^\ell}{\Omega_t} \mathbf{L}_t \tilde{f}_{t+1}\big(\xi_t^\ell\big). \qquad \square$$

A.2. Proof of Theorem 1

We proceed by induction and assume that the claim of the theorem holds for $t \in \mathbb{N}$. To establish (i) for $t+1$, write, using Lemma 11,

$$\begin{aligned}
&\frac{1}{N} \sum_{i=1}^{N} \omega_{t+1}^{i} \big\{\tau_{t+1}^{i} f_{t+1}\big(\xi_{t+1}^{i}\big) + \tilde{f}_{t+1}\big(\xi_{t+1}^{i}\big)\big\} - \phi_t \mathbf{L}_t (\mathbf{T}_{t+1} h_{t+1} f_{t+1} + \tilde{f}_{t+1}) \\
&\quad = \frac{1}{N} \sum_{i=1}^{N} \omega_{t+1}^{i} \big\{\tau_{t+1}^{i} f_{t+1}\big(\xi_{t+1}^{i}\big) + \tilde{f}_{t+1}\big(\xi_{t+1}^{i}\big)\big\} \\
&\qquad - \mathbb{E}\big[\omega_{t+1}^{1} \big\{\tau_{t+1}^{1} f_{t+1}\big(\xi_{t+1}^{1}\big) + \tilde{f}_{t+1}\big(\xi_{t+1}^{1}\big)\big\} \mid \tilde{\mathcal{F}}_t^N\big] \\
&\qquad + \sum_{i=1}^{N} \frac{\omega_t^i}{\Omega_t} \big\{\tau_t^i \mathbf{L}_t f_{t+1}\big(\xi_t^i\big) + \mathbf{L}_t (\tilde{h}_t f_{t+1} + \tilde{f}_{t+1})\big(\xi_t^i\big)\big\} \\
&\qquad - \phi_t \big\{\mathbf{T}_t h_t \mathbf{L}_t f_{t+1} + \mathbf{L}_t (\tilde{h}_t f_{t+1} + \tilde{f}_{t+1})\big\}.
\end{aligned}$$

Since the functions $\mathbf{L}_t f_{t+1}$ and $\mathbf{L}_t (\tilde{h}_t f_{t+1} + \tilde{f}_{t+1})$ belong to $\mathsf{F}_{\mathrm{b}}(\mathcal{X})$, the induction hypothesis (ii) implies that for all $\varepsilon \in \mathbb{R}_+^*$,

$$\begin{aligned}
&\mathbb{P}\Bigg(\Bigg|\sum_{i=1}^{N} \frac{\omega_t^i}{\Omega_t} \big\{\tau_t^i \mathbf{L}_t f_{t+1}\big(\xi_t^i\big) + \mathbf{L}_t (\tilde{h}_t f_{t+1} + \tilde{f}_{t+1})\big(\xi_t^i\big)\big\} \\
&\qquad - \phi_t \big\{\mathbf{T}_t h_t \mathbf{L}_t f_{t+1} + \mathbf{L}_t (\tilde{h}_t f_{t+1} + \tilde{f}_{t+1})\big\}\Bigg| \geq \varepsilon\Bigg) \\
&\quad \leq c_t \exp\big(-\tilde{c}_t N \varepsilon^2\big).
\end{aligned}$$

In addition, by Lemma 12, $\{\omega_{t+1}^i (\tau_{t+1}^i f_{t+1}(\xi_{t+1}^i) + \tilde{f}_{t+1}(\xi_{t+1}^i))\}_{i=1}^N$ are conditionally i.i.d. given $\tilde{\mathcal{F}}_t^N$; thus, since for all $i \in [\![1, N]\!]$,

$$\begin{aligned}
&\big|\omega_{t+1}^{i} \big\{\tau_{t+1}^{i} f_{t+1}\big(\xi_{t+1}^{i}\big) + \tilde{f}_{t+1}\big(\xi_{t+1}^{i}\big)\big\}\big| \\
&\quad \leq \|g_{t+1}\|_\infty \big(\|h_{t+1}\|_\infty \|f_{t+1}\|_\infty + \|\tilde{f}_{t+1}\|_\infty\big) < \infty,
\end{aligned}$$

the conditional Hoeffding inequality provides constants $(d, \tilde{d}) \in (\mathbb{R}_+^*)^2$ such that

$$\mathbb{P}\left(\left|\frac{1}{N}\sum_{i=1}^{N}\omega_{t+1}^{i}\left\{\tau_{t+1}^{i} f_{t+1}\left(\xi_{t+1}^{i}\right)+\tilde{f}_{t+1}\left(\xi_{t+1}^{i}\right)\right\}\right.\right.$$
$$\left.\left.-\mathbb{E}\left[\omega_{t+1}^{1}\left\{\tau_{t+1}^{1} f_{t+1}\left(\xi_{t+1}^{1}\right)+\tilde{f}_{t+1}\left(\xi_{t+1}^{1}\right)\right\} \mid \tilde{\mathcal{F}}_{t}^{N}\right]\right| \geq \varepsilon\right)$$
$$\leq d\exp\left(-\tilde{d}N\varepsilon^{2}\right).$$

This establishes (i).

The inequality (ii) for the self-normalized estimator is an immediate consequence of (i) and the generalized Hoeffding inequality in [11], Lemma 4.

Finally, we conclude the proof by checking that the result is straightforwardly true for the base case $t = 0$, since $\mathbf{T}_0 h_0 = 0$, $\tau_0^i = 0$ for all $i \in [\![1, N]\!]$, and the weighted sample $\{(\xi_0^i, \omega_0^i)\}_{i=1}^N$ (targeting ϕ_0) is generated by standard importance sampling.

A.3. Proof of Theorem 3

We proceed by induction and suppose that the claim of the theorem holds true for some $t \in \mathbb{N}$. Thus, pick $(f_{t+1}, \tilde{f}_{t+1}) \in \mathsf{F}_\mathrm{b}(\mathcal{X})^2$ and assume first that $\phi_{t+1}(\mathbf{T}_{t+1} h_{t+1} f_{t+1} + \tilde{f}_{t+1}) = 0$. Write

$$\sqrt{N}\sum_{i=1}^{N}\frac{\omega_t^i}{\Omega_t}\left\{\tau_{t+1}^{i} f_{t+1}\left(\xi_{t+1}^{i}\right)+\tilde{f}_{t+1}\left(\xi_{t+1}^{i}\right)\right\}$$
$$= N\Omega_t^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\omega_{t+1}^{i}\left\{\tau_{t+1}^{i} f_{t+1}\left(\xi_{t+1}^{i}\right)+\tilde{f}_{t+1}\left(\xi_{t+1}^{i}\right)\right\}\right.$$
$$\left.-\sum_{\ell=1}^{N}\frac{\omega_t^\ell}{\Omega_t}\left\{\tau_t^\ell \mathbf{L}_t f_{t+1}\left(\xi_{t+1}^{\ell}\right)+\mathbf{L}_t(\tilde{h}_t f_{t+1}+\tilde{f}_{t+1})\left(\xi_{t+1}^{\ell}\right)\right\}\right)$$
$$+ N\Omega_t^{-1}\sqrt{N}\sum_{\ell=1}^{N}\frac{\omega_t^\ell}{\Omega_t}\left\{\tau_t^\ell \mathbf{L}_t f_{t+1}\left(\xi_{t+1}^{\ell}\right)+\mathbf{L}_t(\tilde{h}_t f_{t+1}+\tilde{f}_{t+1})\left(\xi_{t+1}^{\ell}\right)\right\},$$

where, by Theorem 1, $N\Omega_t^{-1}$ tends to $(\phi_t \mathbf{L}_t \mathbb{1}_\mathsf{X})^{-1}$ in probability. In order to establish the weak convergence of the first term, we will apply Theorem 16 to the triangular array

$$\upsilon_N^i := \frac{1}{\tilde{N}\sqrt{N}}\sum_{j=1}^{\tilde{N}}\tilde{\upsilon}_N\left(J_{t+1}^{(i,j)}, \xi_{t+1}^{i}\right) \qquad \left(i \in [\![1, N]\!], N \in \mathbb{N}^*\right),$$

where

$$\tilde{\upsilon}_N(j,x) := g_{t+1}(x)\big\{\big(\tau_t^j + \tilde{h}_t\big(\xi_t^j, x\big)\big) f_{t+1}(x) + \tilde{f}_{t+1}(x)\big\} \\ - \sum_{\ell=1}^{N} \frac{\omega_t^\ell}{\Omega_t}\big\{\tau_t^\ell \mathbf{L}_t f_{t+1}\big(\xi_t^\ell\big) + \mathbf{L}_t(\tilde{h}_t f_{t+1} + \tilde{f}_{t+1})\big(\xi_t^\ell\big)\big\} \tag{A.2} \\ \big(x \in \mathsf{X}, j \in [\![1, \tilde{N}]\!], N \in \mathbb{N}^*\big),$$

furnished with the filtration $\{\tilde{\mathcal{F}}_t^N\}_{N\in\mathbb{N}^*}$. Note that for all $i \in [\![1, N]\!]$, $\mathbb{E}[\upsilon_N^i \mid \tilde{\mathcal{F}}_t^N] = 0$ (by Lemma 12) and $|\upsilon_N^i| \le 2\|g_{t+1}\|_\infty(\|h_{t+1}\|_\infty\|f_{t+1}\|_\infty + \|\tilde{f}_{t+1}\|_\infty)/\sqrt{N}$. To check the condition (B1) in Theorem 16, write, using, first, that $\{\upsilon_N^i\}_{i=1}$ are conditionally i.i.d. given $\tilde{\mathcal{F}}_t^N$ and, second, that the backward indices $\{J_{t+1}^{(i,j)}\}_{j=1}^{\tilde{N}}$ are, for all $i \in [\![1, N]\!]$, i.i.d. conditionally on $\tilde{\mathcal{F}}_t^N$ and ξ_{t+1}^i,

$$\sum_{i=1}^{N} \mathbb{E}\big[\big(\upsilon_N^i\big)^2 \mid \tilde{\mathcal{F}}_t^N\big] = \tilde{N}^{-2}\mathbb{E}\left[\left(\sum_{j=1}^{\tilde{N}} \tilde{\upsilon}_N\big(J_{t+1}^{(1,j)}, \xi_{t+1}^1\big)\right)^2 \,\middle|\, \tilde{\mathcal{F}}_t^N\right] \\ = \tilde{N}^{-1}\mathbb{E}\big[\mathbb{E}\big[\tilde{\upsilon}_N^2\big(J_{t+1}^{(1,1)}, \xi_{t+1}^1\big) \mid \mathcal{F}_{t+1}^N\big] \mid \tilde{\mathcal{F}}_t^N\big] \tag{A.3} \\ + \tilde{N}^{-1}(\tilde{N}-1)\mathbb{E}\big[\mathbb{E}^2\big[\tilde{\upsilon}_N\big(J_{t+1}^{(1,1)}, \xi_{t+1}^1\big) \mid \mathcal{F}_{t+1}^N\big] \mid \tilde{\mathcal{F}}_t^N\big]. \tag{A.4}$$

We treat separately the two terms (A.3) and (A.4). Concerning (A.3),

$$\mathbb{E}\big[\mathbb{E}\big[\tilde{\upsilon}_N^2\big(J_{t+1}^{(1,1)}, \xi_{t+1}^1\big) \mid \mathcal{F}_{t+1}^N\big] \mid \tilde{\mathcal{F}}_t^N\big] \\ = \mathbb{E}\left[\sum_{\ell=1}^{N} \tilde{\upsilon}_N^2\big(\ell, \xi_{t+1}^1\big) \frac{\omega_t^\ell q(\xi_t^\ell, \xi_{t+1}^1)}{\sum_{\ell'=1}^{N} \omega_t^{\ell'} q(\xi_t^{\ell'}, \xi_{t+1}^1)} \,\middle|\, \tilde{\mathcal{F}}_t^N\right] \\ = \sum_{i=1}^{N} \frac{\omega_t^i}{\Omega_t} \int q\big(\xi_t^i, x\big) \sum_{\ell=1}^{N} \tilde{\upsilon}_N^2(\ell, x) \frac{\omega_t^\ell q(\xi_t^\ell, x)}{\sum_{\ell'=1}^{N} \omega_t^{\ell'} q(\xi_t^{\ell'}, x)} \mu(\mathrm{d}x) \\ = \sum_{\ell=1}^{N} \frac{\omega_t^\ell}{\Omega_t} \int q\big(\xi_t^\ell, x\big) \tilde{\upsilon}_N^2(\ell, x) \mu(\mathrm{d}x).$$

Now, using the definition (A.2),

$$\sum_{\ell=1}^{N} \frac{\omega_t^\ell}{\Omega_t} \int q\big(\xi_t^\ell, x\big) \tilde{\upsilon}_N^2(\ell, x) \mu(\mathrm{d}x) \\ = \sum_{\ell=1}^{N} \frac{\omega_t^\ell}{\Omega_t} \big(\tau_t^\ell\big)^2 \mathbf{L}_t\big(g_{t+1} f_{t+1}^2\big)\big(\xi_t^\ell\big)$$

$$+\sum_{\ell=1}^{N}\frac{\omega_t^\ell}{\Omega_t}\big(\tau_t^\ell 2\mathbf{L}_t\{g_{t+1}f_{t+1}(\tilde{h}_t f_{t+1}+\tilde{f}_{t+1})\}(\xi_t^\ell)+\mathbf{L}_t\{g_{t+1}(\tilde{h}_t f_{t+1}+\tilde{f}_{t+1})^2\}(\xi_t^\ell)\big)$$
$$-\left(\sum_{\ell=1}^{N}\frac{\omega_t^\ell}{\Omega_t}\{\tau_t^\ell\mathbf{L}_t f_{t+1}(\xi_t^\ell)+\mathbf{L}_t(\tilde{h}_t f_{t+1}+\tilde{f}_{t+1})(\xi_t^\ell)\}\right)^2.$$

In the previous expression, by Lemma 13,

$$\sum_{\ell=1}^{N}\frac{\omega_t^\ell}{\Omega_t}(\tau_t^\ell)^2\mathbf{L}_t(g_{t+1}f_{t+1}^2)(\xi_t^\ell)\xrightarrow{\mathbb{P}}\phi_t\{\mathbf{T}_t^2 h_t\mathbf{L}_t(g_{t+1}f_{t+1}^2)\}+\eta_t\{\mathbf{L}_t(g_{t+1}f_{t+1}^2)\},$$

where the functional η_t is defined in (A.9), and, by Corollary 2 and Lemma 11 (recalling that $\phi_t\mathbf{L}_t(\mathbf{T}_{t+1}f_{t+1}+\tilde{f}_{t+1})=0$ by assumption),

$$\sum_{\ell=1}^{N}\frac{\omega_t^\ell}{\Omega_t}\big(\tau_t^\ell 2\mathbf{L}_t\{g_{t+1}f_{t+1}(\tilde{h}_t f_{t+1}+\tilde{f}_{t+1})\}(\xi_t^\ell)+\mathbf{L}_t\{g_{t+1}(\tilde{h}_t f_{t+1}+\tilde{f}_{t+1})^2\}(\xi_t^\ell)\big)$$
$$\xrightarrow{\mathbb{P}}2\phi_t\big(\mathbf{T}_t h_t\mathbf{L}_t\{g_{t+1}f_{t+1}(\tilde{h}_t f_{t+1}+\tilde{f}_{t+1})\}\big)+\phi_t\mathbf{L}_t\{g_{t+1}(\tilde{h}_t f_{t+1}+\tilde{f}_{t+1})^2\},$$
$$\sum_{\ell=1}^{N}\frac{\omega_t^\ell}{\Omega_t}\{\tau_t^\ell\mathbf{L}_t f_{t+1}(\xi_t^\ell)+\mathbf{L}_t(\tilde{h}_t f_{t+1}+\tilde{f}_{t+1})(\xi_t^\ell)\}\xrightarrow{\mathbb{P}}0. \tag{A.5}$$

We hence conclude that

$$\begin{aligned}(\text{A.3})\xrightarrow{\mathbb{P}}\ &\tilde{N}^{-1}\big(\phi_t\{\mathbf{T}_t^2 h_t\mathbf{L}_t(g_{t+1}f_{t+1}^2)\}+\eta_t\{\mathbf{L}_t(g_{t+1}f_{t+1}^2)\}\\ &+2\phi_t\big(\mathbf{T}_t h_t\mathbf{L}_t\{g_{t+1}f_{t+1}(\tilde{h}_t f_{t+1}+\tilde{f}_{t+1})\}\big)+\phi_t\mathbf{L}_t\{g_{t+1}(\tilde{h}_t f_{t+1}+\tilde{f}_{t+1})^2\}\big)\\ =\ &\tilde{N}^{-1}\big(\phi_t\mathbf{L}_t\big(g_{t+1}\{(\mathbf{T}_t h_t+\tilde{h}_t)f_{t+1}+\tilde{f}_{t+1}\}^2\big)+\eta_t\{\mathbf{L}_t(g_{t+1}f_{t+1}^2)\}\big).\end{aligned} \tag{A.6}$$

We turn to (A.4) and write

$$\begin{aligned}&\mathbb{E}\big[\mathbb{E}^2\big[\tilde{\upsilon}_N(J_{t+1}^{(1,1)},\xi_{t+1}^1)\mid\mathcal{F}_{t+1}^N\big]\mid\tilde{\mathcal{F}}_t^N\big]\\ &=\mathbb{E}\left[\left(\sum_{\ell=1}^{N}\tilde{\upsilon}_N(\ell,\xi_{t+1}^1)\frac{\omega_t^\ell q(\xi_t^\ell,\xi_{t+1}^1)}{\sum_{\ell'=1}^N\omega_t^{\ell'}q(\xi_t^{\ell'},\xi_{t+1}^1)}\right)^2\,\middle|\,\tilde{\mathcal{F}}_t^N\right]\\ &=\sum_{i=1}^{N}\frac{\omega_t^i}{\Omega_t}\int q(\xi_t^i,x)\left(\sum_{\ell=1}^{N}\tilde{\upsilon}_N(\ell,x)\frac{\omega_t^\ell q(\xi_t^\ell,x)}{\sum_{\ell'=1}^N\omega_t^{\ell'}q(\xi_t^{\ell'},x)}\right)^2\mu(\mathrm{d}x),\end{aligned}$$

where we note that the right-hand side can, by (A.2), be written as $\phi_t^N \mathbf{Q} \varphi_N$, with

$$\varphi_N(x) := \Bigg(g_{t+1}(x) f_{t+1}(x) \sum_{\ell=1}^{N} \frac{\omega_t^\ell q(\xi_t^\ell, x)}{\sum_{\ell'=1}^{N} \omega_t^{\ell'} q(\xi_t^{\ell'}, x)} \big\{ \tau_t^\ell + \tilde{h}_t(\xi_t^\ell, x) \big\} + g_{t+1}(x) \tilde{f}_{t+1}(x) \\ - \sum_{\ell'=1}^{N} \frac{\omega_t^{\ell'}}{\Omega_t} \big\{ \tau_t^{\ell'} \mathbf{L}_t f_{t+1}(\xi_t^{\ell'}) + \mathbf{L}_t (\tilde{h}_t f_{t+1} + \tilde{f}_{t+1})(\xi_t^{\ell'}) \big\} \Bigg)^2 \qquad (x \in \mathsf{X}).$$

Note that $\|\varphi_N\|_\infty \leq 4 \|g_{t+1}\|_\infty^2 (\|h_{t+1}\|_\infty \|f_{t+1}\|_\infty + \|\tilde{f}_{t+1}\|_\infty)^2$ for all $N \in \mathbb{N}$. Moreover, by Corollary 2 (and, in particular, the implication (A.5)) it holds, for all $x \in \mathsf{X}$, $\mathbb{P}$-a.s.,

$$\begin{aligned} \varphi_N(x) &\to g_{t+1}^2(x) \Bigg(f_{t+1}(x) \frac{\int q(\tilde{x}, x) \{\mathbf{T}_t h_t(\tilde{x}) + \tilde{h}_t(\tilde{x}, x)\} \phi_t(d\tilde{x})}{\int q(\tilde{x}, x) \phi_t(d\tilde{x})} + \tilde{f}_{t+1}(x) \Bigg)^2 \\ &= g_{t+1}^2(x) \big\{ f_{t+1}(x) \overleftarrow{\mathbf{Q}}_{\phi_t} (\mathbf{T}_t h_t + \tilde{h}_t)(x) + \tilde{f}_{t+1}(x) \big\}^2 \\ &= g_{t+1}^2(x) \big\{ f_{t+1}(x) \mathbf{T}_{t+1} h_{t+1}(x) + \tilde{f}_{t+1}(x) \big\}^2. \end{aligned}$$

Thus, under Assumption 1 we may apply Lemma 14, yielding

$$\phi_t^N \mathbf{Q} \varphi_N \xrightarrow{\mathbb{P}} \phi_t \mathbf{L}_t \big\{ g_{t+1} (f_{t+1} \mathbf{T}_{t+1} h_{t+1} + \tilde{f}_{t+1})^2 \big\},$$

and we may hence conclude that

$$(\text{A.4}) \xrightarrow{\mathbb{P}} \tilde{N}^{-1} (\tilde{N} - 1) \phi_t \mathbf{L}_t \big\{ g_{t+1} (f_{t+1} \mathbf{T}_{t+1} h_{t+1} + \tilde{f}_{t+1})^2 \big\}.$$

Finally, by combining this limit with (A.6) we obtain, using the identity

$$\begin{aligned} &\phi_t \mathbf{L}_t \big(g_{t+1} \big\{ (\mathbf{T}_t h_t + \tilde{h}_t) f_{t+1} + \tilde{f}_{t+1} \big\}^2 \big) - \phi_t \mathbf{L}_t \big\{ g_{t+1} (f_{t+1} \mathbf{T}_{t+1} h_{t+1} + \tilde{f}_{t+1})^2 \big\} \\ &\quad = \phi_t \otimes \mathbf{Q} \big(g_{t+1}^2 \big\{ (\mathbf{T}_t h_t + \tilde{h}_t) f_{t+1} + \tilde{f}_{t+1} \big\}^2 - g_{t+1}^2 \{ f_{t+1} \mathbf{T}_{t+1} h_{t+1} + \tilde{f}_{t+1} \}^2 \big) \\ &\quad = \phi_t \mathbf{Q} \otimes \overleftarrow{\mathbf{Q}}_{\phi_t} \big(g_{t+1}^2 \big\{ (\mathbf{T}_t h_t + \tilde{h}_t) f_{t+1} + \tilde{f}_{t+1} \big\}^2 - g_{t+1}^2 \{ f_{t+1} \mathbf{T}_{t+1} h_{t+1} + \tilde{f}_{t+1} \}^2 \big) \\ &\quad = \phi_t \mathbf{Q} \otimes \overleftarrow{\mathbf{Q}}_{\phi_t} \big\{ g_{t+1}^2 f_{t+1}^2 (\mathbf{T}_t h_t + \tilde{h}_t - \mathbf{T}_{t+1} h_{t+1})^2 \big\} \\ &\quad = \phi_t \mathbf{L}_t \overleftarrow{\mathbf{Q}}_{\phi_t} \big\{ g_{t+1} f_{t+1}^2 (\mathbf{T}_t h_t + \tilde{h}_t - \mathbf{T}_{t+1} h_{t+1})^2 \big\}, \end{aligned}$$

the convergence

$$\begin{aligned} &\sum_{i=1}^{N} \mathbb{E}\big[(\upsilon_N^i)^2 \mid \tilde{\mathcal{F}}_t^N \big] \\ &\quad \xrightarrow{\mathbb{P}} \phi_t \mathbf{L}_t \big\{ g_{t+1} \big(f_{t+1} \mathbf{T}_{t+1} h_{t+1} + \tilde{f}_{t+1} \big)^2 \big\} \\ &\qquad + \tilde{N}^{-1} \big(\phi_t \mathbf{L}_t \overleftarrow{\mathbf{Q}}_{\phi_t} \big\{ g_{t+1} f_{t+1}^2 (\mathbf{T}_t h_t + \tilde{h}_t - \mathbf{T}_{t+1} h_{t+1})^2 \big\} + \eta_t \big\{ \mathbf{L}_t \big(g_{t+1} f_{t+1}^2 \big) \big\} \big), \end{aligned} \tag{A.7}$$

which verifies (B1). In order to check also the condition (B2), write, for $\varepsilon \in \mathbb{R}_+^*$,

$$\sum_{i=1}^{N} \mathbb{E}\big[\big(\upsilon_N^i\big)^2 \mathbb{1}_{\{|\upsilon_N^i| \geq \varepsilon\}} \mid \tilde{\mathcal{F}}_t^N\big] \leq 4\|g_{t+1}\|_\infty^2 \big(\|h_{t+1}\|_\infty \|f_{t+1}\|_\infty + \|\tilde{f}_{t+1}\|_\infty\big)^2 \\ \times \mathbb{1}_{\{2\|g_{t+1}\|_\infty(\|h_{t+1}\|_\infty \|f_{t+1}\|_\infty + \|\tilde{f}_{t+1}\|_\infty) \geq \varepsilon\sqrt{N}\}},$$

where the indicator function on the right-hand side is zero for N large enough. This shows the condition (B2). Hence, for general $(f_{t+1}, \tilde{f}_{t+1})$ (by just replacing $\tilde{f}_{t+1}$ by $\tilde{f}_{t+1} - \phi_{t+1}(\mathbf{T}_{t+1} h_{t+1} f_{t+1} + \tilde{f}_{t+1})$), by Theorem 16, [38], Lemma A.5, and Slutsky's lemma,

$$\sqrt{N} \sum_{i=1}^{N} \frac{\omega_{t+1}^i}{\Omega_{t+1}} \big\{\tau_{t+1}^i f_{t+1}\big(\xi_{t+1}^i\big) + \tilde{f}_{t+1}\big(\xi_{t+1}^i\big) - \phi_{t+1}(\mathbf{T}_{t+1} h_{t+1} f_{t+1} + \tilde{f}_{t+1})\big\} \\ \xrightarrow{\mathcal{D}} \sigma_{t+1}\langle f_{t+1}, \tilde{f}_{t+1}\rangle(h) Z,$$

where Z is a standard Gaussian variable and

$$\begin{aligned} &\sigma_{t+1}^2\langle f_{t+1}, \tilde{f}_{t+1}\rangle(h_{t+1}) \\ &\quad := \frac{\phi_t \mathbf{L}_t (g_{t+1}\{f_{t+1} \mathbf{T}_{t+1} h_{t+1} + \tilde{f}_{t+1} - \phi_{t+1}(\mathbf{T}_{t+1} h_{t+1} f_{t+1} + \tilde{f}_{t+1})\}^2)}{(\phi_t \mathbf{L}_t \mathbb{1}_{\mathsf{X}})^2} \\ &\qquad + \sum_{\ell=0}^{t} \tilde{N}^{\ell-(t+1)} \frac{\phi_\ell \mathbf{L}_\ell \{\overleftarrow{\mathbf{Q}}_{\phi_\ell} (\mathbf{T}_\ell h_\ell + \tilde{h}_\ell - \mathbf{T}_{\ell+1} h_{\ell+1})^2 \mathbf{L}_{\ell+1} \cdots \mathbf{L}_t (g_{t+1} f_{t+1}^2)\}}{(\phi_\ell \mathbf{L}_\ell \cdots \mathbf{L}_{t-1} \mathbb{1}_{\mathsf{X}})(\phi_t \mathbf{L}_t \mathbb{1}_{\mathsf{X}})^2} \\ &\qquad + \frac{\sigma_t^2 (\mathbf{L}_t f_{t+1}, \mathbf{L}_t \{\tilde{h}_{t+1} f_{t+1} + \tilde{f}_{t+1} - \phi_{t+1}(\mathbf{T}_{t+1} h_{t+1} f_{t+1} + \tilde{f}_{t+1})\})}{(\phi_t \mathbf{L}_t \mathbb{1}_{\mathsf{X}})^2}. \end{aligned} \tag{A.8}$$

We now apply the induction hypothesis to the last term. For this purpose, note that, by Lemma 11,

$$\begin{aligned} &h_t \mathbf{L}_t f_{t+1} + \mathbf{L}_t \big\{\tilde{h}_{t+1} f_{t+1} + \tilde{f}_{t+1} - \phi_{t+1}(\mathbf{T}_{t+1} h_{t+1} f_{t+1} + \tilde{f}_{t+1})\big\} \\ &\qquad - \phi_t\big(h_t \mathbf{L}_t f_{t+1} + \mathbf{L}_t \big\{\tilde{h}_{t+1} f_{t+1} + \tilde{f}_{t+1} - \phi_{t+1}(\mathbf{T}_{t+1} h_{t+1} f_{t+1} + \tilde{f}_{t+1})\big\}\big) \\ &\quad = \mathbf{L}_t \big\{h_{t+1} f_{t+1} + \tilde{f}_{t+1} - \phi_{t+1}(\mathbf{T}_{t+1} h_{t+1} f_{t+1} + \tilde{f}_{t+1})\big\}, \end{aligned}$$

yielding, for all $\mathbb{N} \ni s < t$,

$$\begin{aligned} &\tilde{\mathbf{D}}_{s+1,t}\big(h_t \mathbf{L}_t f_{t+1} + \mathbf{L}_t \big\{\tilde{h}_{t+1} f_{t+1} + \tilde{f}_{t+1} - \phi_{t+1}(\mathbf{T}_{t+1} h_{t+1} f_{t+1} + \tilde{f}_{t+1})\big\}\big) \\ &\quad = \mathbf{D}_{s+1,t} \mathbf{L}_t \big\{h_{t+1} f_{t+1} + \tilde{f}_{t+1} - \phi_{t+1}(\mathbf{T}_{t+1} h_{t+1} f_{t+1} + \tilde{f}_{t+1})\big\} \\ &\quad = \tilde{\mathbf{D}}_{s+1,t+1}(h_{t+1} f_{t+1} + \tilde{f}_{t+1}). \end{aligned}$$

We may hence conclude that

$$\begin{aligned}
&\frac{\sigma_t^2(\mathbf{L}_t f_{t+1}, \mathbf{L}_t\{\tilde{h}_{t+1} f_{t+1} + \tilde{f}_{t+1} - \phi_{t+1}(\mathbf{T}_{t+1} h_{t+1} f_{t+1} + \tilde{f}_{t+1})\})}{(\phi_t \mathbf{L}_t \mathbb{1}_\mathsf{X})^2} \\
&\quad = \sum_{s=0}^{t-1} \frac{\phi_s \mathbf{L}_s\{g_{s+1} \tilde{\mathbf{D}}^2_{s+1,t+1}(h_{t+1} f_{t+1} + \tilde{f}_{t+1})\}}{(\phi_s \mathbf{L}_s \cdots \mathbf{L}_t \mathbb{1}_\mathsf{X})^2} \\
&\qquad + \sum_{s=0}^{t-1} \sum_{\ell=0}^{s} \tilde{N}^{\ell-(s+1)} \\
&\qquad \times \frac{\phi_\ell \mathbf{L}_\ell\{\overleftarrow{\mathbf{Q}}_{\phi_\ell}(\mathbf{T}_\ell h_\ell + \tilde{h}_\ell - \mathbf{T}_{\ell+1} h_{\ell+1})^2 \mathbf{L}_{\ell+1} \cdots \mathbf{L}_s(g_{s+1}\{\mathbf{L}_{s+1} \cdots \mathbf{L}_t f_{t+1}\}^2)\}}{(\phi_\ell \mathbf{L}_\ell \cdots \mathbf{L}_{s-1} \mathbb{1}_\mathsf{X})(\phi_s \mathbf{L}_s \cdots \mathbf{L}_t \mathbb{1}_\mathsf{X})^2}.
\end{aligned}$$

Finally, we complete the induction step by noting that

$$\begin{aligned}
&\frac{\phi_t \mathbf{L}_t(g_{t+1}\{f_{t+1} \mathbf{T}_{t+1} h_{t+1} + \tilde{f}_{t+1} - \phi_{t+1}(\mathbf{T}_{t+1} h_{t+1} f_{t+1} + \tilde{f}_{t+1})\}^2)}{(\phi_t \mathbf{L}_t \mathbb{1}_\mathsf{X})^2} \\
&\quad = \frac{\phi_t \mathbf{L}_t\{g_{t+1} \tilde{\mathbf{D}}^2_{t+1,t+1}(h_{t+1} f_{t+1} + \tilde{f}_{t+1})\}}{(\phi_t \mathbf{L}_t \mathbb{1}_\mathsf{X})^2}.
\end{aligned}$$

It remains to check the base case; however, letting, in (A.8), $t = 0$ and $\sigma_0^2 \equiv 0$ (as $\mathbf{T}_0 h_0 = 0$ and $\tau_0^i = 0$ for all $i \in [\![1, N]\!]$) yields

$$\begin{aligned}
\sigma_1^2\langle f_1, \tilde{f}_1\rangle(h_1) &= \frac{\phi_0 \mathbf{L}_0(g_1\{f_1 \mathbf{T}_1 h_1 + \tilde{f}_1 - \phi_1(\mathbf{T}_1 h_1 f_1 + \tilde{f}_1)\}^2)}{(\phi_0 \mathbf{L}_0 \mathbb{1}_\mathsf{X})^2} \\
&\quad + \tilde{N}^{-1} \frac{\phi_0 \mathbf{L}_0\{\overleftarrow{\mathbf{Q}}_{\phi_0}(\tilde{h}_0 - \mathbf{T}_1 h_1)^2 g_1 f_1^2\}}{(\phi_0 \mathbf{L}_0 \mathbb{1}_\mathsf{X})^2} \\
&= \frac{\phi_0 \mathbf{L}_0\{g_1 \tilde{\mathbf{D}}_{1,1}(h_1 f_1 + \tilde{f}_1)^2\}}{(\phi_0 \mathbf{L}_0 \mathbb{1}_\mathsf{X})^2} + \tilde{N}^{-1} \frac{\phi_0 \mathbf{L}_0\{\overleftarrow{\mathbf{Q}}_{\phi_0}(\tilde{h}_0 - \mathbf{T}_1 h_1)^2 g_1 f_1^2\}}{(\phi_0 \mathbf{L}_0 \mathbb{1}_\mathsf{X})^2}
\end{aligned}$$

which is, under the standard convention that $\mathbf{L}_m \mathbf{L}_n = \mathrm{id}$ if $m > n$, in agreement with (3.2). This completes the proof.

Lemma 13. *Let Assumption 1 hold. Then for all $t \in \mathbb{N}$, $f_t \in \mathsf{F}_\mathrm{b}(\mathcal{X})$, and $\tilde{N} \in \mathbb{N}^*$,*

$$\sum_{i=1}^{N} \frac{\omega_t^i}{\Omega_t} (\tau_t^i)^2 f_t(\xi_t^i) \xrightarrow{\mathbb{P}} \phi_t(\mathbf{T}_t^2 h_t f_t) + \eta_t(f_t),$$

where

$$\eta_t(f_t) := \sum_{\ell=0}^{t-1} \tilde{N}^{\ell-t} \frac{\phi_\ell \mathbf{L}_\ell \{\overleftarrow{\mathbf{Q}}_{\phi_\ell}(\mathbf{T}_\ell h_\ell + \tilde{h}_\ell - \mathbf{T}_{\ell+1} h_{\ell+1})^2 \mathbf{L}_{\ell+1} \cdots \mathbf{L}_{t-1} f_t\}}{\phi_\ell \mathbf{L}_\ell \cdots \mathbf{L}_{t-1} \mathbb{1}_{\mathsf{X}}}. \tag{A.9}$$

Proof. Again, we proceed by induction. First, the base case $t = 0$ is trivially true since $\mathbf{T}_0 h_0 = 0$, $\tau_0^i = 0$ for all $i \in [\![1, N]\!]$, and $\sum_{\ell=0}^{-1} = 0$ by convention. We now assume that the claim of the lemma holds true for some $t \in \mathbb{N}$. Since Corollary 2 implies that $N^{-1}\Omega_{t+1} \stackrel{\mathbb{P}}{\longrightarrow} \phi_t \mathbf{L}_t \mathbb{1}_{\mathsf{X}}$ it is enough to study the convergence of $N^{-1}\sum_{i=1}^N \omega_{t+1}^i (\tau_{t+1}^i)^2 f_{t+1}(\xi_{t+1}^i)$. For this purpose, we will apply Theorem 15 to the triangular array

$$\upsilon_N^i := N^{-1}\omega_{t+1}^i \big(\tau_{t+1}^i\big)^2 f_{t+1}\big(\xi_{t+1}^i\big) \qquad \big(i \in [\![1, N]\!], N \in \mathbb{N}^*\big)$$

furnished with the filtration $\{\tilde{\mathcal{F}}_t^N\}_{N\in\mathbb{N}^*}$. Note that $|\upsilon_N^i| \le \|g_{t+1}\|_\infty \|h_{t+1}\|_\infty^2 \|f_{t+1}\|_\infty / N$ for all $i \in [\![1, N]\!]$ and $N \in \mathbb{N}^*$. In addition, using, first that $\{\upsilon_N^i\}_{i=1}^N$ are conditionally i.i.d. given $\tilde{\mathcal{F}}_t^N$ and, second, that for all $i \in [\![1, N]\!]$, the backward indices $\{J_{t+1}^{(i,j)}\}_{j=1}^{\tilde{N}}$ are conditionally i.i.d. given $\tilde{\mathcal{F}}_t^N$ and ξ_{t+1}^i,

$$\begin{aligned}
&\sum_{i=1}^N \mathbb{E}\big[\upsilon_N^i \mid \tilde{\mathcal{F}}_t^N\big] \\
&\quad = \mathbb{E}\big[\omega_{t+1}^1 \big(\tau_{t+1}^1\big)^2 f_{t+1}\big(\xi_{t+1}^1\big) \mid \tilde{\mathcal{F}}_t^N\big] \\
&\quad = \tilde{N}^{-1}\mathbb{E}\big[\omega_{t+1}^1 f_{t+1}\big(\xi_{t+1}^1\big)\mathbb{E}\big[\big(\tau_t^{J_{t+1}^{(1,1)}} + \tilde{h}_t\big(\xi_t^{J_{t+1}^{(1,1)}}, \xi_{t+1}^1\big)\big)^2 \mid \mathcal{F}_{t+1}^N\big] \mid \tilde{\mathcal{F}}_t^N\big] && \text{(A.10)} \\
&\qquad + \tilde{N}^{-1}(\tilde{N} - 1)\mathbb{E}\big[\omega_{t+1}^1 f_{t+1}\big(\xi_{t+1}^1\big)\mathbb{E}^2\big[\tau_t^{J_{t+1}^{(1,1)}} + \tilde{h}_t\big(\xi_t^{J_{t+1}^{(1,1)}}, \xi_{t+1}^1\big) \mid \mathcal{F}_{t+1}^N\big] \mid \tilde{\mathcal{F}}_t^N\big]. && \text{(A.11)}
\end{aligned}$$

We treat separately the two terms (A.10) and (A.11). First,

$$\begin{aligned}
\text{(A.10)} &= \tilde{N}^{-1}\mathbb{E}\Bigg[\omega_{t+1}^1 f_{t+1}\big(\xi_{t+1}^1\big) \sum_{\ell=1}^N \frac{\omega_t^\ell q(\xi_t^\ell, \xi_{t+1}^1)}{\sum_{\ell'=1}^N \omega_t^{\ell'} q(\xi_t^{\ell'}, \xi_{t+1}^1)} \big\{\tau_t^\ell + \tilde{h}_t\big(\xi_t^\ell, \xi_{t+1}^1\big)\big\}^2 \,\Big|\, \tilde{\mathcal{F}}_t^N\Bigg] \\
&= \tilde{N}^{-1}\sum_{i=1}^N \frac{\omega_t^i}{\Omega_t} \int q\big(\xi_t^i, x\big) g_{t+1}(x) f_{t+1}(x) \\
&\qquad \times \sum_{\ell=1}^N \frac{\omega_t^\ell q(\xi_t^\ell, x)}{\sum_{\ell'=1}^N \omega_t^{\ell'} q(\xi_t^{\ell'}, x)} \big\{\tau_t^\ell + \tilde{h}_t\big(\xi_t^\ell, x\big)\big\}^2 \mu(\mathrm{d}x) \\
&= \tilde{N}^{-1}\sum_{\ell=1}^N \frac{\omega_t^\ell}{\Omega_t} \int q\big(\xi_t^\ell, x\big) g_{t+1}(x) f_{t+1}(x) \big\{\tau_t^\ell + \tilde{h}_t\big(\xi_t^\ell, x\big)\big\}^2 \mu(\mathrm{d}x).
\end{aligned}$$

Using Corollary 2 and the induction hypothesis, we obtain the limits

$$\sum_{\ell=1}^{N} \frac{\omega_t^\ell}{\Omega_t} (\tau_t^\ell)^2 \mathbf{L}_t(\xi_t^\ell, f_{t+1}) \xrightarrow{\mathbb{P}} \phi_t(\mathbf{T}_t^2 h_t \mathbf{L}_t f_{t+1}) + \eta_t(\mathbf{L}_t f_{t+1}),$$

$$\sum_{\ell=1}^{N} \frac{\omega_t^\ell}{\Omega_t} \tau_t^\ell \mathbf{L}_t(\xi_t^\ell, \tilde{h}_t f_{t+1}) \xrightarrow{\mathbb{P}} \phi_t\{\mathbf{T}_t h_t \mathbf{L}_t(\tilde{h}_t f_{t+1})\},$$

$$\sum_{\ell=1}^{N} \frac{\omega_t^\ell}{\Omega_t} \mathbf{L}_t(\xi_t^\ell, \tilde{h}_t^2 f_{t+1}) \xrightarrow{\mathbb{P}} \phi_t \mathbf{L}_t(\tilde{h}_t^2 f_{t+1}),$$

which yield

$$\begin{aligned}(\text{A.10}) &\xrightarrow{\mathbb{P}} \tilde{N}^{-1}\big(\phi_t(\mathbf{T}_t^2 h_t \mathbf{L}_t f_{t+1}) + \eta_t(\mathbf{L}_t f_{t+1}) + 2\phi_t\{\mathbf{T}_t h_t \mathbf{L}_t(\tilde{h}_t f_{t+1})\} + \phi_t \mathbf{L}_t(\tilde{h}_t^2 f_{t+1})\big) \\ &= \tilde{N}^{-1}\big(\phi_t \mathbf{L}_t\{(\mathbf{T}_t h_t + \tilde{h}_t)^2 f_{t+1}\} + \eta_t(\mathbf{L}_t f_{t+1})\big).\end{aligned}$$

We turn to the second term (A.11) and equate the same with $\tilde{N}^{-1}(\tilde{N}-1)\phi_t^N \mathbf{Q}\varphi_N$, where

$$\varphi_N(x) := g_{t+1}(x) f_{t+1}(x) \left(\sum_{\ell=1}^{N} \frac{\omega_t^\ell q(\xi_t^\ell, x)}{\sum_{\ell'=1}^{N} \omega_t^{\ell'} q(\xi_t^{\ell'}, x)} \{\tau_t^\ell + \tilde{h}_t(\xi_t^\ell, x)\} \right)^2 \qquad (x \in \mathsf{X}).$$

Since $\|\varphi_N\|_\infty \le \|g_{t+1}\|_\infty \|f_{t+1}\|_\infty \|h_{t+1}\|_\infty$ for all $N \in \mathbb{N}$, and, by Corollary 2, for all $x \in \mathsf{X}$, $\mathbb{P}$-a.s.,

$$\begin{aligned}\varphi_N(x) &\to g_{t+1}(x) f_{t+1}(x) \left(\frac{\int q(\tilde{x}, x)\{\mathbf{T}_t h_t(\tilde{x}) + \tilde{h}_t(\tilde{x}, x)\}\phi_t(\mathrm{d}\tilde{x})}{\int q(\tilde{x}, x)\phi_t(\mathrm{d}\tilde{x})} \right)^2 \\ &= g_{t+1}(x) f_{t+1}(x) \overleftarrow{\mathbf{Q}}_{\phi_t}^2 (\mathbf{T}_t h_t + \tilde{h}_t)(x) \\ &= g_{t+1}(x) f_{t+1}(x) \mathbf{T}_{t+1}^2 h_{t+1}(x),\end{aligned}$$

we may, under Assumption 1, apply Lemma 14, yielding

$$\phi_t^N \mathbf{Q}\varphi_N \xrightarrow{\mathbb{P}} \phi_t \mathbf{Q}(g_{t+1} f_{t+1} \mathbf{T}_{t+1}^2 h_{t+1}) = \phi_t \mathbf{L}_t(\mathbf{T}_{t+1}^2 h_{t+1} f_{t+1}).$$

Consequently,

$$\begin{aligned}\sum_{i=1}^{N} \mathbb{E}[\upsilon_N^i \mid \tilde{\mathcal{F}}_t^N] &\xrightarrow{\mathbb{P}} \phi_t \mathbf{L}_t(\mathbf{T}_{t+1}^2 h_{t+1} f_{t+1}) \\ &\quad + \tilde{N}^{-1}\big(\phi_t \mathbf{L}_t\{(\mathbf{T}_t h_t + \tilde{h}_t)^2 f_{t+1}\} - \phi_t \mathbf{L}_t(\mathbf{T}_{t+1}^2 h_{t+1} f_{t+1}) + \eta_t(\mathbf{L}_t f_{t+1})\big).\end{aligned} \qquad \text{(A.12)}$$

In order to show that $\sum_{i=1}^N \upsilon_N^i$ has the same limit (A.12) in probability we use Theorem 15. Condition (A1) is easily checked by reusing (A.12) with f_{t+1} replaced by $|f_{t+1}|$. In order to check (A2), we simply note that for all $\varepsilon \in \mathbb{R}_+^*$,

$$\sum_{i=1}^N \mathbb{E}\big[\big|\upsilon_N^i\big|\mathbb{1}_{\{|\upsilon_N^i|\geq\varepsilon\}} \mid \tilde{\mathcal{F}}_t^N\big] \leq \|g_{t+1}\|_\infty \|h_{t+1}\|_\infty^2 \|f_{t+1}\|_\infty \mathbb{1}_{\{\|g_{t+1}\|_\infty \|h_{t+1}\|_\infty^2 \|f_{t+1}\|_\infty \geq \varepsilon N\}},$$

where the right-hand side is zero for N large enough. Thus, Theorem 15 applies and since, by reversibility,

$$\begin{aligned}
&\phi_t \mathbf{L}_t\big\{(\mathbf{T}_t h_t + \tilde{h}_t)^2 f_{t+1}\big\} - \phi_t \mathbf{L}_t\big(\mathbf{T}_{t+1}^2 h_{t+1} f_{t+1}\big) \\
&\quad = \phi_t \otimes \mathbf{Q}\big(\big\{(\mathbf{T}_t h_t + \tilde{h}_t)^2 - \mathbf{T}_{t+1}^2 h_{t+1}\big\} g_{t+1} f_{t+1}\big) \\
&\quad = \phi_t \mathbf{Q} \otimes \overleftarrow{\mathbf{Q}}_{\phi_t}\big\{(\mathbf{T}_t h_t + \tilde{h}_t - \mathbf{T}_{t+1} h_{t+1})^2 g_{t+1} f_{t+1}\big\} \\
&\quad = \phi_t \mathbf{L}_t\big\{\overleftarrow{\mathbf{Q}}_{\phi_t}(\mathbf{T}_t h_t + \tilde{h}_t - \mathbf{T}_{t+1} h_{t+1})^2 f_{t+1}\big\},
\end{aligned}$$

Slutsky's lemma implies

$$\begin{aligned}
&\sum_{i=1}^N \frac{\omega_{t+1}^i}{\Omega_{t+1}} \big(\tau_{t+1}^i\big)^2 f_{t+1}\big(\xi_{t+1}^i\big) \\
&\quad = N \Omega_{t+1}^{-1} \sum_{i=1}^N \upsilon_N^i \\
&\quad \xrightarrow{\mathbb{P}} \phi_{t+1}\big(\mathbf{T}_{t+1}^2 h_{t+1} f_{t+1}\big) \\
&\qquad + \frac{1}{\tilde{N}(\phi_t \mathbf{L}_t \mathbb{1}_{\mathsf{X}})} \big(\phi_t \mathbf{L}_t\big\{\overleftarrow{\mathbf{Q}}_{\phi_t}(\mathbf{T}_t h_t + \tilde{h}_t - \mathbf{T}_{t+1} h_{t+1})^2 f_{t+1}\big\} + \eta_t(\mathbf{L}_t f_{t+1})\big).
\end{aligned}$$

We may now conclude the proof by noting, using the induction hypothesis, the identity

$$(\phi_\ell \mathbf{L}_\ell \cdots \mathbf{L}_{t-1} \mathbb{1}_{\mathsf{X}})(\phi_t \mathbf{L}_t \mathbb{1}_{\mathsf{X}}) = \phi_\ell \mathbf{L}_\ell \cdots \mathbf{L}_t \mathbb{1}_{\mathsf{X}},$$

and the convention $\mathbf{L}_{t+1}\mathbf{L}_t = \mathrm{id}$, that

$$\begin{aligned}
&\frac{1}{\tilde{N}(\phi_t \mathbf{L}_t \mathbb{1}_{\mathsf{X}})} \big(\phi_t \mathbf{L}_t\big\{\overleftarrow{\mathbf{Q}}_{\phi_t}(\mathbf{T}_t h_t + \tilde{h}_t - \mathbf{T}_{t+1} h_{t+1})^2 f_{t+1}\big\} + \eta_t(\mathbf{L}_t f_{t+1})\big) \\
&\quad = \sum_{\ell=0}^t \tilde{N}^{\ell-(t+1)} \frac{\phi_\ell \mathbf{L}_\ell\{\overleftarrow{\mathbf{Q}}_{\phi_\ell}(\mathbf{T}_\ell h_\ell + \tilde{h}_\ell - \mathbf{T}_{\ell+1} h_{\ell+1})^2 \mathbf{L}_{\ell+1} \cdots \mathbf{L}_t f_{t+1}\}}{\phi_\ell \mathbf{L}_\ell \cdots \mathbf{L}_t \mathbb{1}_{\mathsf{X}}}.
\end{aligned}$$

□

The following lemma formalizes an argument used in the proof of [11], Theorem 8.

Lemma 14. *Let Assumption 1 hold. Let $\mathbf{\Psi}$ be a possibly unnormalized transition kernel on $(\mathsf{X}, \mathcal{X})$ having transition density $\psi \in \mathsf{F}_\mathrm{b}(\mathcal{X}^2)$ with respect to some reference measure λ. Moreover, let $\{\varphi_N\}_{N \in \mathbb{N}^*}$ be a sequence of functions in $\mathsf{F}_\mathrm{b}(\mathcal{X})$ for which*:

(i) *there exists $\varphi \in \mathsf{F}_\mathrm{b}(\mathcal{X})$ such that for all $x \in \mathsf{X}$, $\varphi_N(x) \to \varphi(x)$, $\mathbb{P}$-a.s., and*
(ii) *there exists $|\varphi|_\infty \in \mathbb{R}_+^*$ such that $\|\varphi_N\|_\infty \leq |\varphi|_\infty$ for all $N \in \mathbb{N}^*$.*

Then for all $t \in \mathbb{N}$, $\phi_t^N \mathbf{\Psi} \varphi_N \xrightarrow{\mathbb{P}} \phi_t \mathbf{\Psi} \varphi$.

Proof. Since, by Corollary 2, $\phi_t^N \mathbf{\Psi} \varphi \xrightarrow{\mathbb{P}} \phi_t \mathbf{\Psi} \varphi$, it is enough to establish that

$$\phi_t^N \mathbf{\Psi} \varphi_N \xrightarrow{\mathbb{P}} \phi_t^N \mathbf{\Psi} \varphi.$$

For this purpose, set

$$a_N(x) := \left|\varphi_N(x) - \varphi(x)\right| \int \psi(\tilde{x}, x) \phi_t^N(\mathrm{d}\tilde{x}),$$
$$\tilde{a}_N(x) := \int \psi(\tilde{x}, x) \phi_t^N(\mathrm{d}\tilde{x}) \qquad \big(N \in \mathbb{N}^*, x \in \mathsf{X}\big).$$

Since $|\phi_t^N \mathbf{\Psi} \varphi_N - \phi_t^N \mathbf{\Psi} \varphi| \leq \lambda a_N$ it is, by Markov's inequality, enough to show that $\mathbb{E}[\lambda a_N]$ tends to zero as N tends to infinity. However, by Fubini's theorem,

$$\lim_{N \to \infty} \mathbb{E}[\lambda a_N] = \lim_{N \to \infty} \int \mathbb{E}\big[a_N(x)\big] \lambda(\mathrm{d}x) = 0,$$

where the last equality is a consequence of the generalized Lebesgue dominated convergence theorem provided that:

(i) $\lim_{N \to \infty} \mathbb{E}[a_N(x)] = 0$ for all $x \in \mathsf{X}$,
(ii) there exists $c \in \mathbb{R}_+^*$ such that $\mathbb{E}[a_N(x)] \leq c \mathbb{E}[\tilde{a}_N(x)]$ for all $x \in \mathsf{X}$,
(iii) $\lim_{N \to \infty} \int \mathbb{E}[\tilde{a}_N(x)] \lambda(\mathrm{d}x) = \int \lim_{N \to \infty} \mathbb{E}[\tilde{a}_N(x)] \lambda(\mathrm{d}x)$.

Here, (i) is implied by Corollary 2 and, as $|a_N(x)| \leq \|\psi\|_\infty (\|\varphi\|_\infty + |\varphi|_\infty)$ for all $x \in \mathsf{X}$, the standard dominated convergence theorem. Moreover, (ii) is satisfied with $c = \|\varphi\|_\infty + |\varphi|_\infty$. Finally, to check (iii), notice that

$$\begin{aligned} \lim_{N \to \infty} \int \mathbb{E}\big[\tilde{a}_N(x)\big] \lambda(\mathrm{d}x) &\overset{\text{(a)}}{=} \lim_{N \to \infty} \mathbb{E}\big[\phi_t^N \mathbf{\Psi} \mathbb{1}_\mathsf{X}\big] \overset{\text{(b)}}{=} \phi_t \mathbf{\Psi} \mathbb{1}_\mathsf{X} \\ &\overset{\text{(c)}}{=} \iint \psi(\tilde{x}, x) \phi_t(\mathrm{d}\tilde{x}) \lambda(\mathrm{d}x) \overset{\text{(d)}}{=} \int \lim_{N \to \infty} \mathbb{E}\big[\tilde{a}_N(x)\big] \lambda(\mathrm{d}x), \end{aligned}$$

where (a) and (c) follow by Fubini's theorem and (b) and (d) are obtained from Corollary 2 and the standard dominated convergence theorem (as $\mathbf{\Psi} \mathbb{1}_\mathsf{X} \in \mathsf{F}_\mathrm{b}(\mathcal{X})$ and $\psi \in \mathsf{F}_\mathrm{b}(\mathcal{X}^2)$ by assumption). This completes the proof. □

A.4. Proof of Proposition 7

By [18], Lemma 1,

$$\|\tilde{\mathbf{D}}_{s+1,t}h_t\|_\infty \le \|\mathbf{L}_{s+1}\cdots\mathbf{L}_t\mathbb{1}_{\mathsf{X}}\|_\infty \sum_{\ell=0}^{t-1} \varrho^{\max\{s-\ell+2,\ell-s-3,0\}} \operatorname{osc}(\tilde{h}_\ell), \tag{A.13}$$

and, consequently,

$$\sum_{s=0}^{t-1} \frac{\phi_s\mathbf{L}_s(g_{s+1}\tilde{\mathbf{D}}^2_{s+1,t}h_t)}{(\phi_s\mathbf{L}_s\cdots\mathbf{L}_{t-1}\mathbb{1}_{\mathsf{X}})^2} \le |\tilde{h}|^2_\infty \sum_{s=0}^{t-1} \frac{(\phi_s\mathbf{L}_s g_{s+1})\|\mathbf{L}_{s+1}\cdots\mathbf{L}_t\mathbb{1}_{\mathsf{X}}\|^2_\infty}{(\phi_s\mathbf{L}_s\cdots\mathbf{L}_{t-1}\mathbb{1}_{\mathsf{X}})^2}\left(\sum_{\ell=0}^{t-1}\varrho^{\max\{s-\ell+2,\ell-s-3,0\}}\right)^2.$$

Now, under Assumption 2, for all $x \in \mathsf{X}$,

$$\underline{\varepsilon}\mu(g_{s+2}\mathbf{L}_{s+2}\cdots\mathbf{L}_t\mathbb{1}_{\mathsf{X}}) \le \mathbf{L}_{s+1}\cdots\mathbf{L}_t\mathbb{1}_{\mathsf{X}}(x) \le \bar{\varepsilon}\mu(g_{s+2}\mathbf{L}_{s+2}\cdots\mathbf{L}_t\mathbb{1}_{\mathsf{X}}) \tag{A.14}$$

implying that

$$\frac{(\phi_s\mathbf{L}_s g_{s+1})\|\mathbf{L}_{s+1}\cdots\mathbf{L}_t\mathbb{1}_{\mathsf{X}}\|^2_\infty}{(\phi_s\mathbf{L}_s\cdots\mathbf{L}_{t-1}\mathbb{1}_{\mathsf{X}})^2} = \frac{(\phi_{s+1}g_{s+1})\|\mathbf{L}_{s+1}\cdots\mathbf{L}_t\mathbb{1}_{\mathsf{X}}\|^2_\infty}{(\phi_s\mathbf{L}_s\mathbb{1}_{\mathsf{X}})(\phi_{s+1}\mathbf{L}_{s+1}\cdots\mathbf{L}_{t-1}\mathbb{1}_{\mathsf{X}})^2} \le \frac{\bar{\delta}}{\underline{\delta}}\left(\frac{\bar{\varepsilon}}{\underline{\varepsilon}}\right)^2 = \frac{\bar{\delta}}{\underline{\delta}(1-\varrho)^2}.$$

Moreover, as

$$\begin{aligned}\sum_{s=0}^{t-1}\left(\sum_{\ell=0}^{t-1}\varrho^{\max\{s-\ell+2,\ell-s-3,0\}}\right)^2 &= \sum_{s=0}^{t-4}\left(\frac{2-\varrho^{s+3}-\varrho^{t-s-3}}{1-\varrho}\right)^2 + \sum_{s=t-3}^{t-1}\left(\rho^{s+2}\frac{1-\rho^{-t}}{1-\rho^{-1}}\right)\\ &= \frac{4t}{(1-\varrho)^2} + o(t),\end{aligned}$$

we conclude that

$$\limsup_{t\to\infty}\frac{1}{t}\sum_{s=0}^{t-1}\frac{\phi_s\mathbf{L}_s(g_{s+1}\tilde{\mathbf{D}}^2_{s+1,t}h_t)}{(\phi_s\mathbf{L}_s\cdots\mathbf{L}_{t-1}\mathbb{1}_{\mathsf{X}})^2} \le |\tilde{h}|^2_\infty\frac{4\bar{\delta}}{\underline{\delta}(1-\varrho)^4}.$$

A.5. Proof of Theorem 8

The first term of $\sigma_t^2(h_t)$ is the asymptotic variance of the FFBSm algorithm, which is, by Proposition 7, bounded by $4|\tilde{h}|^2_\infty\bar{\delta}/\{\underline{\delta}(1-\varrho)^4\}$. To treat the second term, we bound, using

(A.14),

$$\begin{aligned}
&\frac{\mathbf{L}_{\ell+1}\cdots\mathbf{L}_s(g_{s+1}\{\mathbf{L}_{s+1}\cdots\mathbf{L}_{t-1}\mathbb{1}_\mathsf{X}\}^2)}{(\phi_{\ell+1}\mathbf{L}_{\ell+1}\cdots\mathbf{L}_{s-1}\mathbb{1}_\mathsf{X})(\phi_s\mathbf{L}_s\cdots\mathbf{L}_{t-1}\mathbb{1}_\mathsf{X})^2} \\
&\quad\le \frac{\|\mathbf{L}_{\ell+1}\cdots\mathbf{L}_{s-1}\mathbb{1}_\mathsf{X}\|_\infty\bar{\delta}\|\mathbf{L}_{s+1}\cdots\mathbf{L}_{t-1}\mathbb{1}_\mathsf{X}\|_\infty^2}{(\phi_{\ell+1}\mathbf{L}_{\ell+1}\cdots\mathbf{L}_{s-1}\mathbb{1}_\mathsf{X})(\phi_s\mathbf{L}_s\mathbb{1}_\mathsf{X})(\phi_{s+1}\mathbf{L}_{s+1}\cdots\mathbf{L}_{t-1}\mathbb{1}_\mathsf{X})^2} \qquad (A.15)\\
&\quad\le \frac{\bar{\delta}}{\underline{\delta}(1-\varrho)^3}.
\end{aligned}$$

Moreover, since $\mathbf{T}_{\ell+1}h_{\ell+1} = \overleftarrow{\mathbf{Q}}_{\phi_\ell}(\mathbf{T}_\ell h_\ell + \tilde{h}_\ell)$ and $\mathbf{T}_\ell h_\ell = \mathbf{D}_{\ell,\ell}h_\ell$, we obtain, by reusing (A.13),

$$\begin{aligned}
\|\mathbf{T}_\ell h_\ell + \tilde{h}_\ell - \mathbf{T}_{\ell+1}h_{\ell+1}\|_\infty &\le \operatorname{osc}(\mathbf{T}_\ell h_\ell) + \operatorname{osc}(\tilde{h}_\ell) \le 4\|\tilde{\mathbf{D}}_{\ell,\ell}h_\ell\|_\infty + \operatorname{osc}(\tilde{h}_\ell) \\
&\le |\tilde{h}|_\infty\left(\frac{4\bar{\delta}\varrho^2}{1-\varrho} + 1\right).
\end{aligned}$$

Thus,

$$\begin{aligned}
&\frac{1}{t}\sum_{s=0}^{t-1}\sum_{\ell=0}^{s}\tilde{N}^{\ell-(s+1)} \\
&\qquad\times \frac{\phi_{\ell+1}\{\overleftarrow{\mathbf{Q}}_{\phi_\ell}(\mathbf{T}_\ell h_\ell + \tilde{h}_\ell - \mathbf{T}_{\ell+1}h_{\ell+1})^2\mathbf{L}_{\ell+1}\cdots\mathbf{L}_s(g_{s+1}\{\mathbf{L}_{s+1}\cdots\mathbf{L}_{t-1}f_t\}^2)\}}{(\phi_{\ell+1}\mathbf{L}_{\ell+1}\cdots\mathbf{L}_{s-1}\mathbb{1}_\mathsf{X})(\phi_s\mathbf{L}_s\cdots\mathbf{L}_{t-1}\mathbb{1}_\mathsf{X})^2} \\
&\quad\le |\tilde{h}|_\infty^2\frac{\bar{\delta}}{\underline{\delta}(1-\varrho)^3}\left(\frac{4\bar{\delta}\varrho^2}{1-\varrho}+1\right)^2\frac{1}{t}\sum_{s=0}^{t-1}\sum_{\ell=0}^{s}\tilde{N}^{\ell-(s+1)},
\end{aligned}$$

and since

$$\lim_{s\to\infty}\sum_{\ell=0}^{s}\tilde{N}^{\ell-(s+1)} = (\tilde{N}-1)^{-1}$$

we may conclude the proof by taking the Cesàro mean.

A.6. Proof of Theorem 9

In the case of marginal smoothing, [11], Theorem 12, provides, for $t \ge \hat{s}$ (since the variance vanishes for $t < \hat{s}$, the result holds trivially true in this case), the time uniform bound

$$\tilde{\sigma}_t^2(h_t) \le \operatorname{osc}^2(\tilde{h}_{\hat{s}})\frac{\bar{\delta}^2(1+\varrho^2)}{(1+\varrho)(1-\varrho)^3}$$

and hence, since all terms are zero except $\tilde{h}_{\hat{s}}$, it is enough to bound the quantity

$$\sum_{s=\hat{s}}^{t-1} \sum_{\ell=\hat{s}}^{s} \tilde{N}^{\ell-(s+1)} \times \frac{\phi_{\ell+1}\{\overleftarrow{\mathbf{Q}}_{\phi_\ell}(\mathbf{T}_\ell h_\ell + \tilde{h}_\ell - \mathbf{T}_{\ell+1} h_{\ell+1})^2 \mathbf{L}_{\ell+1} \cdots \mathbf{L}_s (g_{s+1}\{\mathbf{L}_{s+1} \cdots \mathbf{L}_{t-1} f_t\}^2)\}}{(\phi_{\ell+1}\mathbf{L}_{\ell+1} \cdots \mathbf{L}_{s-1} \mathbb{1}_\mathsf{X})(\phi_s \mathbf{L}_s \cdots \mathbf{L}_{t-1} \mathbb{1}_\mathsf{X})^2} \tag{A.16}$$

(where $\tilde{h}_\ell = 0$ for $\ell > \hat{s}$). In addition, by [11], Lemma 10, for all $\ell \geq \hat{s}$,

$$\|\tilde{\mathbf{D}}_{\ell,\ell} h_\ell\|_\infty \leq \varrho^{\ell-\hat{s}} \operatorname{osc}(\tilde{h}_{\hat{s}}),$$

yielding

$$\|\mathbf{T}_\ell h_\ell + \tilde{h}_\ell - \mathbf{T}_{\ell+1} h_{\ell+1}\|_\infty \leq 2\varrho^{\ell-\hat{s}} \operatorname{osc}(\tilde{h}_{\hat{s}}).$$

By combining this with (A.15) we obtain, via standard operations on geometric sums,

$$\begin{aligned}
(\text{A.16}) &\leq 4 \operatorname{osc}^2(\tilde{h}_{\hat{s}}) \frac{\bar{\delta}}{\underline{\delta}(1-\varrho)^3} \sum_{s=\hat{s}}^{t-1} \sum_{\ell=\hat{s}}^{s} \tilde{N}^{\ell-(s+1)} \varrho^{2(\ell-\hat{s})} \\
&= 4 \operatorname{osc}^2(\tilde{h}_{\hat{s}}) \frac{\bar{\delta}}{\underline{\delta}(1-\varrho)^3} \\
&\quad \times \begin{cases} \dfrac{1}{(1-\tilde{N}\varrho^2)} \left(\dfrac{1-\tilde{N}^{-(t-\hat{s})}}{\tilde{N}-1} - \varrho^2 \dfrac{1-\varrho^{2(t-\hat{s})}}{1-\varrho^2} \right) & \text{if } \tilde{N}\varrho^2 \neq 1, \\ \dfrac{1}{\tilde{N}-1} \left(2 - \tilde{N}^{-(t-\hat{s})} - (t-\hat{s})\tilde{N}^{-(t-\hat{s})+1} - \tilde{N}^{-(t-\hat{s})}\right) & \text{if } \tilde{N}\varrho^2 = 1, \end{cases}
\end{aligned}$$

and hence, letting t tend to infinity,

$$(\text{A.16}) \leq 4 \operatorname{osc}^2(\tilde{h}_{\hat{s}}) \frac{\bar{\delta}}{\underline{\delta}(1-\varrho)^3} \times \begin{cases} \dfrac{1}{(\tilde{N}-1)(1-\varrho^2)} & \text{if } \tilde{N}\varrho^2 \neq 1, \\ \dfrac{2}{\tilde{N}-1} & \text{if } \tilde{N}\varrho^2 = 1, \end{cases}$$

which concludes the proof.

Appendix B: Technical results

B.1. Conditional limit theorems for triangular arrays of dependent random variables

We first recall two results, obtained in [12] (but reformulated slightly here for our purposes), which are essential for the developments of the present paper.

Theorem 15 ([12]). *Let* $(\Omega, \mathcal{A}, \{\mathcal{F}_N\}_{N \in \mathbb{N}^*}, \mathbb{P})$ *be a filtered probability space. In addition, let* $\{\upsilon_N^i\}_{i=1}^N$, $N \in \mathbb{N}^*$, *be a triangular array of random variables on* $(\Omega, \mathcal{A}, \mathbb{P})$ *such that for all* $N \in \mathbb{N}^*$, *the variables* $\{\upsilon_N^i\}_{i=1}^N$ *are conditionally independent given* $\mathcal{F}_N$ *with* $\mathbb{E}[|\upsilon_N^i| \mid \mathcal{F}_N] < \infty$, $\mathbb{P}$-*a.s., for all* $i \in [\![1, N]\!]$. *Moreover, assume that*

(A1)

$$\lim_{\lambda \to \infty} \sup_{N \in \mathbb{N}^*} \mathbb{P}\left(\sum_{i=1}^N \mathbb{E}\left[\left| \upsilon_N^i \right| \mid \mathcal{F}_N \right] \geq \lambda \right) = 0.$$

(A2) *For all* $\varepsilon > 0$, *as* $N \to \infty$,

$$\sum_{i=1}^N \mathbb{E}\left[\left| \upsilon_N^i \right| \mathbb{1}_{\{|\upsilon_N^i| \geq \varepsilon\}} \mid \mathcal{F}_N \right] \xrightarrow{\mathbb{P}} 0.$$

Then, as $N \to \infty$,

$$\max_{m \in [\![1, N]\!]} \left| \sum_{i=1}^m \upsilon_N^i - \sum_{i=1}^m \mathbb{E}\left[\upsilon_N^i \mid \mathcal{F}_N \right] \right| \xrightarrow{\mathbb{P}} 0.$$

Theorem 16 ([12]). *Let the assumptions of Theorem 15 hold with* $\mathbb{E}[(\upsilon_N^i)^2 \mid \mathcal{F}_N] < \infty$, $\mathbb{P}$-*a.s., for all* $i \in [\![1, N]\!]$, *and* (A1) *and* (A2) *replaced by*

(B1) *For some constant* $\varsigma^2 > 0$, *as* $N \to \infty$,

$$\sum_{i=1}^N \left(\mathbb{E}\left[\left(\upsilon_N^i\right)^2 \mid \mathcal{F}_N \right] - \mathbb{E}^2\left[\upsilon_N^i \mid \mathcal{F}_N \right] \right) \xrightarrow{\mathbb{P}} \varsigma^2.$$

(B2) *For all* $\varepsilon > 0$, *as* $N \to \infty$,

$$\sum_{i=1}^N \mathbb{E}\left[\left(\upsilon_N^i\right)^2 \mathbb{1}_{\{|\upsilon_N^i| \geq \varepsilon\}} \mid \mathcal{F}_N \right] \xrightarrow{\mathbb{P}} 0.$$

Then, for all $u \in \mathbb{R}$, *as* $N \to \infty$,

$$\mathbb{E}\left[\exp\left(\mathrm{i} u \sum_{i=1}^N \left\{ \upsilon_N^i - \mathbb{E}\left[\upsilon_N^i \mid \mathcal{F}_N \right] \right\} \right) \,\Big|\, \mathcal{F}_N \right] \xrightarrow{\mathbb{P}} \exp\left(-u^2 \varsigma^2 / 2\right).$$

B.2. An accept–reject-based algorithm for backward sampling

Given two subsequent particle samples $\{(\xi_{s-1}^i, \omega_{s-1}^i)\}_{i=1}^N$ and $\{(\xi_s^i, \omega_s^i)\}_{i=1}^N$, the following algorithm, which is a trivial adjustment of [11], Algorithm 1, simulates the full set $\{J_s^{(i,j)} : (i, j) \in [\![1, N]\!] \times [\![1, \tilde{N}]\!]\}$ of backward indices required for one iteration of PaRIS. The algorithm requires Assumption 1(ii) to hold true.

Algorithm 3 Accept-reject-based backward sampling

Require: Particle samples $\{(\xi_{s-1}^i, \omega_{s-1}^i)\}_{i=1}^N$ and $\{(\xi_s^i, \omega_s^i)\}_{i=1}^N$.

for $j = 1 \to \tilde{N}$ **do**
  set $\mathsf{L} \leftarrow [\![1, N]\!]$;
  while $\mathsf{L} \neq \varnothing$ **do**
    set $n \leftarrow \#\mathsf{L}$;
    draw $(I_1, \ldots, I_N) \sim \mathsf{Pr}(\{\omega_{s-1}^i\}_{i=1}^N)^{\otimes N}$;
    draw $(U_1, \ldots, U_N) \sim \mathsf{U}(0, 1)^{\otimes N}$;
    set $\mathsf{L}_n \leftarrow \varnothing$;
    for $k = 1 \to n$ **do**
      if $U_k \leq q(\xi_{s-1}^{I_k}, \xi_s^{\mathsf{L}(k)})/\bar{\varepsilon}$ **then**
        set $J_s^{(\mathsf{L}(k), j)} \leftarrow I_k$;
      else
        set $\mathsf{L}_n \leftarrow \mathsf{L}_n \cup \{\mathsf{L}(k)\}$;
      end if
    end for
    set $\mathsf{L} \leftarrow \mathsf{L}_n$;
  end while
end for
return $\{J_s^{(i,j)} : (i, j) \in [\![1, N]\!] \times [\![1, \tilde{N}]\!]\}$

Acknowledgements

The authors thank the anonymous referees for insightful comments that improved the presentation of the paper. Jimmy Olsson is supported by the Swedish Research Council, Grant 2011-5577.

References

[1] Baum, L.E., Petrie, T., Soules, G. and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Stat.* **41** 164–171. MR0287613

[2] Cappé, O. (2001). Ten years of HMMs (online bibliography 1989–2000).

[3] Cappé, O. (2011). Online EM algorithm for hidden Markov models. *J. Comput. Graph. Statist.* **20** 728–749. MR2878999

[4] Cappé, O., Godsill, S.J. and Moulines, E. (2007). An overview of existing methods and recent advances in sequential Monte Carlo. *IEEE Proceedings* **95** 899–924.

[5] Cappé, O., Moulines, E. and Rydén, T. (2005). *Inference in Hidden Markov Models*. *Springer Series in Statistics*. New York: Springer. With Randal Douc's contributions to Chapter 9 and Christian P. Robert's to Chapters 6, 7 and 13, with Chapter 14 by Gersende Fort, Philippe Soulier and Moulines, and Chapter 15 by Stéphane Boucheron and Elisabeth Gassiat. MR2159833

[6] Chib, S., Nardari, F. and Shephard, N. (2002). Markov chain Monte Carlo methods for stochastic volatility models. *J. Econometrics* **108** 281–316. MR1894758

[7] Crisan, D. and Heine, K. (2008). Stability of the discrete time filter in terms of the tails of noise distributions. *J. Lond. Math. Soc.* (2) **78** 441–458. MR2439634

[8] Del Moral, P. (2004). *Feynman–Kac Formulae. Genealogical and Interacting Particle Systems with Applications. Probability and Its Applications* (*New York*). New York: Springer. MR2044973

[9] Del Moral, P., Doucet, A. and Singh, S.S. (2010). A backward particle interpretation of Feynman–Kac formulae. *ESAIM Math. Model. Numer. Anal.* **44** 947–975. MR2731399

[10] Del Moral, P. and Guionnet, A. (2001). On the stability of interacting processes with applications to filtering and genetic algorithms. *Ann. Inst. Henri Poincaré Probab. Stat.* **37** 155–194. MR1819122

[11] Douc, R., Garivier, A., Moulines, E. and Olsson, J. (2011). Sequential Monte Carlo smoothing for general state space hidden Markov models. *Ann. Appl. Probab.* **21** 2109–2145. MR2895411

[12] Douc, R. and Moulines, E. (2008). Limit theorems for weighted samples with applications to sequential Monte Carlo methods. *Ann. Statist.* **36** 2344–2376. MR2458190

[13] Douc, R., Moulines, E. and Olsson, J. (2014). Long-term stability of sequential Monte Carlo methods under verifiable conditions. *Ann. Appl. Probab.* **24** 1767–1802. MR3226163

[14] Douc, R., Moulines, E. and Stoffer, D.S. (2014). *Nonlinear Time Series*: *Theory*, *Methods*, *and Applications with R Examples*. *Chapman & Hall/CRC Texts in Statistical Science Series*. Boca Raton, FL: Chapman & Hall/CRC. MR3289095

[15] Doucet, A., de Freitas, N. and Gordon, N. (eds.) (2001). *Sequential Monte Carlo Methods in Practice*. *Statistics for Engineering and Information Science*. New York: Springer. MR1847783

[16] Doucet, A., Godsill, S. and Andrieu, C. (2000). On sequential Monte-Carlo sampling methods for Bayesian filtering. *Stat. Comput.* **10** 197–208.

[17] Doucet, A. and Johansen, A.M. (2009). A tutorial on particle filtering and smoothing: Fifteen years later. In *Oxford Handbook of Nonlinear Filtering*. Oxford: Oxford Univ. Press.

[18] Dubarry, C. and Le Corff, S. (2013). Non-asymptotic deviation inequalities for smoothed additive functionals in nonlinear state-space models. *Bernoulli* **19** 2222–2249. MR3160552

[19] Fearnhead, P., Wyncoll, D. and Tawn, J. (2010). A sequential smoothing algorithm with linear computational cost. *Biometrika* **97** 447–464. MR2650750

[20] Godsill, S.J., Doucet, A. and West, M. (2004). Monte Carlo smoothing for non-linear time series. *J. Amer. Statist. Assoc.* **50** 438–449.

[21] Gordon, N., Salmond, D. and Smith, A.F. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc. F, Radar Signal Process.* **140** 107–113.

[22] Hull, J. and White, A. (1987). The pricing of options on assets with stochastic volatilities. *J. Finance* **42** 281–300.

[23] Hürzeler, M. and Künsch, H.R. (1998). Monte Carlo approximations for general state-space models. *J. Comput. Graph. Statist.* **7** 175–193. MR1649366

[24] Jacob, P.E., Murray, L.M. and Rubenthaler, S. (2013). Path storage in the particle filter. *Stat. Comput.* **25** 487–496.

[25] Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *J. Comput. Graph. Statist.* **5** 1–25. MR1380850

[26] Kitagawa, G. and Sato, S. (2001). Monte Carlo smoothing and self-organising state-space model. In *Sequential Monte Carlo Methods in Practice*. *Stat. Eng. Inf. Sci.* 177–195. New York: Springer. MR1847792

[27] Koski, T. (2001). *Hidden Markov Models for Bioinformatics*. Dordrecht: Kluwer Academic Publishers.

[28] Lindsten, F. and Schön, T.B. (2013). Backward simulation methods for Monte Carlo statistical inference. *Foundations and Trends in Machine Learning* **6** 1–143.

[29] Mongillo, G. and Deneve, S. (2008). Online learning with hidden Markov models. *Neural Comput.* **20** 1706–1716. MR2417104

[30] Olsson, J., Cappé, O., Douc, R. and Moulines, E. (2008). Sequential Monte Carlo smoothing with application to parameter estimation in nonlinear state space models. *Bernoulli* **14** 155–179. MR2401658

[31] Olsson, J. and Ströjby, J. (2010). Convergence of random weight particle filters. Technical report, Lund University.

[32] Olsson, J. and Ströjby, J. (2011). Particle-based likelihood inference in partially observed diffusion processes using generalised Poisson estimators. *Electron*. *J*. *Stat*. **5** 1090–1122. MR2836770

[33] Olsson, J. and Westerborn, J. (2014). Efficient particle-based online smoothing in general hidden Markov models. In *IEEE* 2014 *International Conference on Acoustics*, *Speech*, *and Signal Processing* (*ICASSP* 2014). New York: IEEE.

[34] Pitt, M.K. and Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters. *J*. *Amer*. *Statist*. *Assoc*. **94** 590–599. MR1702328

[35] Poyiadjis, G., Doucet, A. and Singh, S.S. (2011). Particle approximations of the score and observed information matrix in state space models with application to parameter estimation. *Biometrika* **98** 65–80. MR2804210

[36] Rabiner, L.R. and Juang, B.-H. (1993). *Fundamentals of Speech Recognition*. New York: Prentice-Hall.

[37] Rauch, H.E., Tung, F. and Striebel, C.T. (1965). Maximum likelihood estimates of linear dynamic systems. *AIAA J*. **3** 1445–1450. MR0181489

[38] Vergé, C., Del Moral, P., Moulines, E. and Olsson, J. (2014). Convergence properties of weighted particle islands with application to the double bootstrap algorithm. Preprint.

Received December 2014 and revised September 2015